DOI: 10.21276/IJIPSR.2018.06.03.659

International Journal of Innovative Pharmaceutical Sciences and Research
www.ijipsr.com

APPLICATION OF POOLED DNA SEQUENCING TECHNOLOGY FOR THE STUDY OF POPULATION GENOMICS IN CROP IMPROVEMENT PROGRAM

[1]Alemu Tebeje, [2]Aragaw Zemene*

*[1]Department of Plant Biotechnology, College of Agriculture & Veterinary Medicine, Jimma University, **ETHIOPIA***
*[2]Department of Biotechnology, College of Natural & Computational Sciences, Adigrat University, **ETHIOPIA***

Abstract

Over the past several decades, especially through the traditional breeding program, intensive attempts have been made for the improvement of a large number of crop varieties which adjusted to diverse agro-ecologies. However, increasing biotic and abiotic stresses, increasing populations, and sharply reducing natural resources especially water for agricultural purposes, push the breeders and researchers for organizing and developing improved crop varieties with higher yield potential. In combination with developments in agricultural technology, plant breeding has made remarkable progress in improving crop qualities for over a century. Modern techniques of plant biotechnology like pooled DNA sequencing are widely employed in plant breeding and crop improvement. Pooled-DNA sequencing is being used for the acceleration of plant selection and genome sequencing with a short period. Genes of agronomic and scientific importance can be isolated especially by their position on the genetic map by using pooled DNA technologies. In this review, the current status of sequencing development technologies for crop improvements has been clearly discussed. It has also tried to recommend and provide an outlook into the future approaches and most widely used applications in plant breeding of crop plants by present development.

Keywords: Population genomics; Sequencing technology; Whole genome sequence; Bar-coding; Pool sequence; Rare allelic variant.

Corresponding Author:

Aragaw Zemene

Department of Biotechnology,

College of Natural & Computational Sciences,

*Adigrat University, **ETHIOPIA***

***E-mail:** zemenearagaw@gmail.com*

***Phone:** +251-942-857-301*

DOI: 10.21276/IJIPSR.2018.06.03.659

INTRODUCTION

At present, global agriculture is facing a serious threat from climate change, which is predicted to result in reduced productivity. In light with this, an increase in global food demand and the potential impact of climate change have created an urgent need for effective crop improvement programs to deliver higher productivity and resilience to the impact of greater unfavorable environmental stresses in agriculture [20]. Crop genetic improvement must simultaneously address the needs of growing demand (higher productivity), adaptation to climate variability and change (climate resilience) and deliver healthy foods (nutritional security). Effective utilization of available genetic resources by the help of different up-to-date and emerging technologies is a key strategy for enhanced crop improvement programs [20]. Knowledge of a crop genome sequence is fundamental for understanding biochemical and physiological processes that govern crop traits and the way in which they respond to environments – and biotic and abiotic stresses. Pooled-DNA sequencing technology is now an important strategy in the studies of crop genetic diversity and adaptation. It supports the management and utilization of germplasm, to characterize and conserve wild populations, and to discover novel genetic resources of wild crop populations [7]. Sequencing of crops genetic material [19] may reveal genetic diversity in parts of the genome that have limited variation due to genetic bottlenecks associated with domestication or crop improvement [24].

Ongoing advances in pooled DNA sequencing technology provide increasing opportunities to understand plant and crop species at the whole-genome level [14]. Pooled-DNA sequencing of the genomes of these species is in some cases supporting the development of reference genome sequences for the species; in other cases, data providing re-sequencing relative to the domesticated genome is the outcome, while for some species the analysis of transcriptomes is allowing the expressed genome to be defined. The rapid evolution of genome sequencing technologies has resulted in an explosion of genomic information, the sequencing of a vast number of plant genomes, and opportunities to apply this to crop improvement programs [34].

Objectives

This review paper is prepared with the hope to achieve the following objectives

- To define the topic, and create understanding by the scientific community
- To review the current state of population genomics and pooled DNA sequencing methods in crop productivity
- To emphasize the link between this technology and plant biotechnology
- To discuss the future impact and power of pooled sequencing on crop improvement practices

DOI: 10.21276/IJIPSR.2018.06.03.659

- Finally to deliver recommendations on:– future research trends and priorities for funding

History & Current Status of Population Genomics in Crop Improvement

Genomics, as a scientific era, is relatively new. Advances in biology and molecular genetics technology predate the advent of genomics, for example, development of cloning vectors by Paul Berg [22]. But the inception of genomics was coincident with the genesis of the human genome project (HGP), which was conceptualized and endorsed by a US Department of Energy-sponsored meeting in Santa Fe, NM (USA), in 1986. Advances in cloning, robotics, DNA preparation, automation of DNA sequencing, computing, and informatics have led to a democratization of genomics such that producing a genome sequence is now affordable and conceivable for many, if not all, crop genomes [22].

Population genomics is a recently coined term which is equivalent to mean population genetics the large-scale comparison of DNA sequences of populations that study genome-wide effects to improve our understanding on micro-evolution [29]. It is the study of the amount and causes of genome-wide variability in natural populations [8]. It is crucial because only genome-wide effects inform us reliably about population demography and phylogenetic history, whereas locus-specific effects help identify genes that are important for fitness & adaptation [8]. Population genomics has been used to understand the reason for the phenotypic variation within a species. However, since the genetic variation within this species was previously poorly understood due to technological restrictions, population genomics allows us to learn about the species' genetic differences [17]. In the human population, population genomics has been used to study the genetic change since humans began to migrate away from Africa approximately 50,000-100,000 years ago. It has been shown that not only were genes related to fertility and reproduction highly selected for but also that the further humans moved away from Africa, the greater the presence of lactase [25]. Previously genomics was restricted to only the study of a low amount of loci. However, recent advancements in sequencing, large-scale genetic screening and computer storage, and power have allowed for the study of hundreds of thousands of loci from populations [38]. These new high-throughput methods also allow researchers to collect vast amounts of information about crop genetic variation in very short periods of time [38].

Population genomics has been of interest to scientists since Darwin. By drawing on evidence from domesticated species of animals and plants, Darwin succeeded in demonstrating the existence of variability in both quantitative and discrete traits [11] & [12]. No progress in understanding the causes of this variability was made, however, until the rediscovery of

DOI: 10.21276/IJIPSR.2018.06.03.659

Mendelian genetics at the beginning of the 20th century. While confirming Darwin's view that there is plenty of genetic variabilities available for use in evolution, the evidence obtained by these methods left two important questions unanswered [26]. First, how much variation within a natural population is there at an average gene locus? The second question concerned the extent to which the frequencies of variants within populations (other than rare deleterious mutations) are controlled by natural selection. Over the past ten years, the field of population genetics has undergone major renovations because of recent advances in gene sequencing and screening technologies. These technological innovations have allowed scientists to tackle bigger and broader questions related to population trends, and to study genetic variation on a much broader scale than ever before possible with older methods, such as test crosses, random sampling, and field work. Today, discoveries can be facilitated by the ever-expanding field of genomics, which is the use of large databases for the purpose of studying genetic variation on a large scale across different organisms [26].

Modern Techniques of Population Genomics for Crop Improvement

Sequencing technology

By knowing the DNA sequence, the cause of the various diseases can be known. We can determine the sequence responsible for various diseases and can be treated with the help of gene therapy. DNA sequencing can solve a lot of problems and perform a lot of works for human welfare. Some important general applications of DNA sequencing are:

1. To analyze any protein structure and function, we must have the knowledge of its primary structure i.e. its DNA sequence.
2. With its study, we can understand the function of a specific sequence and the sequence responsible for any trait or desire qualities of crop and other organisms.
3. With the help of comparative DNA sequence study, we can detect any mutation.
4. Kinship study, and increases the knowledge working with DNA fingerprinting technology.
5. By knowing the whole genome sequence, Human genome project gets completed.

Whole Genome Sequencing

Whole genome sequencing is a laboratory process that determines the complete DNA sequence of an organism's genome at a single time. This entails sequencing all of an organism's chromosomal DNA as well as DNA contained in the mitochondria and, for plants, in the chloroplast. Sequences up to several megabases in length have been found to be present in individual genomes but absent in the human reference genome.

DOI: 10.21276/IJIPSR.2018.06.03.659

RefSeq is a public database of nucleotide and protein sequences with feature and bibliographic annotation. The RefSeq database is built and distributed by the National Center for Biotechnology Information (NCBI). These sequences may be common in populations, and their absence in the reference genome may indicate rare variants in the genomes of individuals who served as donors for the human genome project. As the reference genome is used in probe design for microarray technology and mapping short reads in next generation sequencing (NGS), this missing sequence could be a source of bias in functional genomics studies and variant analysis [28].

Current GWAS are based on the strategy of linkage disequilibrium (LD) mapping, in which a sufficient number of single nucleotide polymorphism (SNP) markers are selectively genotyped to capture the genetic variation in the whole genome. However, there are two major issues related to the results of GWAS. First, the results only explain a small fraction of the heritability of complex traits. One of the reasons may be that many functional variants, in particular, rare variants, which are not directly genotyped in GWAS, have a weak LD with SNP markers, and hence are missed by GWAS [21] & [30]. Second, the identified associations in GWAS are often inconsistent between different populations. The reason for this may be the varied LD structures between markers and underlying causal variants among populations, resulting in associations can only be observed in specific populations. To overcome this problem, an ideal approach is to directly sequence all the samples in a study **[5]**. However, this is not a feasible option for the traditional sequencing technology, namely Sanger sequencing, which is extremely expensive and time consumption for sequencing thousands of samples required to achieve reasonable statistical power in a typical genetic association study.To reduce the cost of large-scale association studies, one efficient approach is to sequence a large number of individuals together on a single sequence run [5].

DNA Barcoding technology

DNA barcoding is a taxonomic method that uses a short genetic marker in an organism's DNA to identify it as belonging to a particular species [18]. It differs from molecular phylogeny in that the main goal is not to determine patterns of relationship but to identify an unknown sample regarding a pre-existing classification [23]. Bar-coding ligates the DNA fragments of each sample to a short, sample-specific DNA sequence, and then sequences these DNA fragments from multiple subjects in one single sequencing run. In addition to allowing determining individual genotypes, it offers an additional advantage of reduction of sequencing variability **[10]**. However, bar-coding at present has a limit of the multiplexing, and the cost of

DOI: 10.21276/IJIPSR.2018.06.03.659

the individual DNA amplification and sequencing template preparation could be substantial in large scale disease-association studies [10].

Pooled-DNA sequencing technology and its application in crop improvement

DNA Pooling was first used in the genetic study, in a case-control association study of HLA class II, DR and DQ alleles in type I diabetes mellitus [2]. Afterwards, it has been used for linkage studies in plants [32]. for the homozygosity mapping of recessive diseases in inbred populations [39], and for mutation detection [1]. This strategy was also proposed for high-throughput SNP arrays [21]. Many GWAS have used pooling to detect the rare causal SNPs [10] & [36]. However, all these methods make the assumption that all individuals have the same abundance level in the pool. The abundance level for each individual is the fraction of the reads in a pool originated [10].

a) Initial genetic maps consisted of few and sparse markers, many of which were anonymous markers (simple sequence repeats (SSR)) or markers based on restriction fragment length polymorphisms (RFLP). For example, if a phenotype of interest were affected by genetic variation within the SSR1-SSR2 interval, the complete region would be selected with little information about its gene content or allelic variation.
b) Whole genome sequencing of a closely related species enabled projection of gene content onto the target genetic map. These allowed breeders to postulate the presence of specific genes by conserved gene order across species (synteny), although this varies between species and regions.
c) Complete genome sequence of the target species provides breeders with an unprecedented wealth of information that allows them to access and identify variation that is useful for crop improvement. In addition to providing immediate access to gene content, putative gene function and precise genomic positions, the whole genome sequence facilitates the identification of both natural and induced (by TILLING) variation in germplasm collections and copy number variation between varieties. Promoter sequences allow epigenetic states to be surveyed, and expression levels can be monitored in different tissues or environments and specific genetic backgrounds using RNAseq or microarrays. Integration of these layers of information can create gene networks, from which epistasis and target pathways can be identified. Furthermore, re-sequencing of varieties identifies a high density of SNP markers across genomic intervals, which enable genome-wide association studies (GWAS), genomic selection (GS) and more defined marker-assisted selection (MAS) strategies.

DOI: 10.21276/IJIPSR.2018.06.03.659

The most accurate measure of genetic variability within accessions would be to compare the band profiles of large numbers of individuals. However, the large numbers of accessions held in most germplasm collections render this approach impractical. A practical alternative might also be to use pools. Large pooled DNA provides little information on the genetic variability within the accessions being studied. For instance, consider the variation between pools comprised of 10-plant mixes in which a particular allele (x) appears in 20% of the genotypes.

Comparison of Bar coding and Pooled-DNA sequencing

The two techniques are contributing different great activities in genome sequencing and analysis of crops and other organisms in general. Compared to barcoding, simply pooling DNA samples is more cost-effective as it can fully make use of the high depth of sequencing and vastly reduce the efforts of sample preparation for thousands of individuals. The basic idea behind this principle is that DNA from multiple individuals is pooled together into a single DNA mixture, prepared as a single library and sequenced [15].

In this approach, the library preparation cost is reduced because one library is prepared per pool instead of one library per sample. It allows allele frequencies in groups of individuals to be measured using far fewer PCR reactions and genotyping assays than are used when genotyping individuals. With pooling, the sequencing throughput required per individual is much less than what is provided by a single run, and hence it is feasible to sequence multiple individuals together. For example, in a case-control study, the allele frequencies in a sample of 500 cases and 500 controls can be measured from two pooled samples, rather than from 1,000 individual samples, which represents an increase in efficiency of 500-fold [15].

Pooling methods can be split into two categories. The first category puts each individual in only one pool and each pool consist of fixed number of individuals. These types of methods are referred to as non-overlapping pool methods. The second category puts each individual in multiple pools and uses this information to recover each individual's genotype. These methods are referred to as overlapping pool methods. In genome-wide analysis studies, two-stage design and DNA pooling could be used as a cost-efficient strategy to detect genetic variant regions [9].

In the first stage, a fraction of samples are genotyped for all SNPs and a case-control association test for each SNP is then conducted to select the most significant SNPs. In the second stage, the candidate SNPs from the first stage is further evaluated by genotyping. To reduce cost of large-scale association studies in two-stage design,pools of DNA from many individuals have been successfully used in the first stage of the two-stage design [3].

DOI: 10.21276/IJIPSR.2018.06.03.659

MATERIALS & METHOD

DNA-pool construction

Several steps are required to construct pools that contain equal quantities of DNA from individual samples and from which robust PCR results can be obtained. In the first step, DNA concentration can be measured by Ultraviolet (UV) light spectroscopy. However, this approach alone is not sufficiently accurate for measuring concentration unless the DNA samples are all of the high purity, as contaminants can affect UV light absorbance. A quantization step that is based on *fluorimetry* with a DNA-specific dye (such as PicoGreen TM) is therefore recommended in some protocols [4]. Inaccuracies can also arise from the pipetting of small volumes of viscous solutions in which DNA concentration is not homogeneous. This can be avoided by using non-viscous, diluted stock samples to construct the pools. Even samples that seem to be of the same concentration can vary in their ability to be amplified by PCR, so samples need to be checked by PCR to identify those that do not yield a robust product [4].

Quantitative genotyping assays

SNPs confer a base compositional difference at a polymorphic site that can be detected in an amplified PCR fragment. Various strategies have been developed to genotype SNPs, each of which has varying potential for use in the analysis of allele frequencies in DNA pools. In the first general approach, an SNP can be exploited to create PCR fragments of differing size. In the simplest assay of this type, the PCR product is digested with a restriction enzyme endonuclease that cleaves a fragment from one bi-allelic SNP but not the other. Alternatively, modified nucleotides can be included in the PCR reaction that becomes incorporated into the PCR product in allele-specific patterns, and that generate sites that are sensitive or resistant to chemical cleavage. Both methods generate PCR products of differing size that represent specific SNP alleles, which can be detected by conventional electrophoresis on gels or in capillary systems [10] & [42]. In the second approach, primers close to or abutting; the variable SNP sites are used in a primer extension reaction. During extension, specific di-deoxyribonucleotides are incorporated that terminate the reaction in a sequence-specific manner, which results in allele-specific extension products [40] & [6]. The allele-specific extension products can be distinguished by numerous methods - for example, by conventional fluorescent tagging and electrophoretic separation. Alternatively, alleles can be detected by pyrosequencing, in which the extension is coupled to a base-specific light-emission reaction [35] & [45]. Much higher throughput can be obtained by applying extension methods to highly automated and highly parallel platforms. A parallel approach can be achieved using

DOI: 10.21276/IJIPSR.2018.06.03.659

microarrays. In this approach, reactions that are analogous to the fluorescence method can be carried out simultaneously for several loci [16]. To distinguish between loci, each extension primer has a unique identifier tag at its 5′-end, which causes it to bind to pre-defined complementary oligonucleotide sites on a microarray. Because these arrays can carry probes that contain many thousands of such sites, their use allows a highly parallel approach [16]. In a third generic strategy, allele-specific hybridization is used to discriminate between local sequences at the polymorphic site. Several methods are available for this type of analysis. The two most common methods resolve alleles by either allele-specific hybridization of primers in a PCR reaction or allele-specific hybridization to primers anchored to a microarray. For any method of SNP detection to be applied to pooling, each step must be quantitative, from PCR through to signal detection. At present, because each SNP detection assay used depends on an initial PCR step, a common problem is the unbiased representation of allelic products that are present in a DNA pool [4]. Most importantly, we must also here consider problems that are associated with the application of the three main approaches listed above in DNA pooling — cleavage, primer extension, and hybridization. The main problem that affects cleavage-based methods is ensuring that the cleavage reactions are complete. Any tendency towards partial cleavage results in a systematic overestimation of the allele that corresponds to the uncut PCR product. The second method, primer extension, is the SNP genotyping method that the results obtained have been good. However, there are two main potential problems that are associated with its use. The first is that each base is not incorporated into the extension reaction with equal efficiency [41]. The second is that partial self-complementarity at the 3′-end of primers might result in self-annealing, which allows the extension to occur independently of the target template. However, this problem can be identified readily by carrying out a control extension reaction in the absence of a template [41].

Strategies of pooling design

The main idea of pooling is to sequence DNA from several individuals together on a single run. Through the observed number of re-sequencing alleles, the allele frequency can be estimated. The simplest strategy is the naïve-pooling scheme, which is also called disjoint pooling. In thenaïve-pooling scheme, DNA was sequenced from several individuals on a single pool and each pool includes different individuals. It offers insight into allele frequencies but is not able to the identity of an allele carrier. Recently, several strategies of well-chosen pools aiming to identify variant are proposed. In these designs, each individual is tested several times in different pools. This redundancy provides a potential increase in both sensitivity and specificity [6].

DOI: 10.21276/IJIPSR.2018.06.03.659

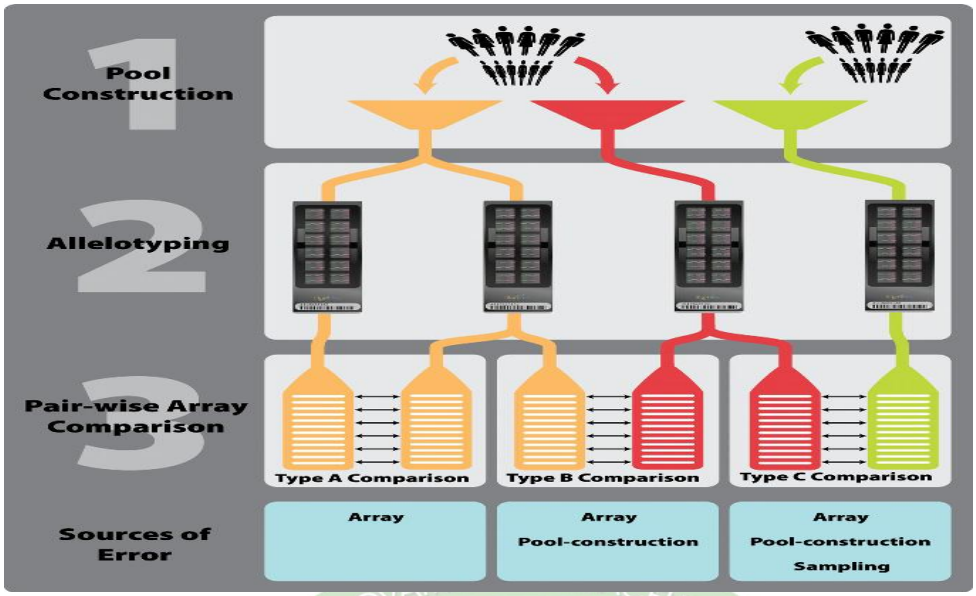

Fig. 1: Overview of the pair-wise array comparison's performed in this study

Step 1 depicts the construction of three DNA pools. The first two pools (orange and red) are constructed using the same DNA samples and are pool-construction replicates. The third pool (green) is constructed using difference DNA samples.

Step 2 indicates allelotyping on Illumina SNP arrays, where the two arrays allelotyping the orange pool are array replicate.

Step 3 shows the three types of pair-wise SNP array comparisons that can be made, along with the sources of error that account for differences in allele frequency estimates on the paired arrays. For Type A comparisons, the arrays being compared were used to allelotype the same DNA pool; hence, the only source of variation is the array. For Type B comparisons, the arrays paired were used to allelotype independently constructed but identical pools; thus, variation may arise due to the array and the pool construction process. For Type C comparisons, the arrays paired were used to allelotype completely independent DNA pools, and variation may be due to the array, pool construction, or binomial sampling (assuming both pools are independent samples from a single population).

For case-control sequencing studies, we propose a design and analysis strategy for DNA pooling that can greatly reduce the cost of sequencing and also allow covariate adjustment for SNP-disease association. Our method includes three steps:

1) **Pool creation:** Case and control samples will be grouped according to the similarity of their characteristics (e.g., age, sex, ethnicity or principal components, etc.). Samples with similar characteristics will be pooled for sequencing.
2) **Genotype Imputation:** DNA pools will be created with the approximately equal amount from each sample. Sequencing will be carried out on the pooled DNA samples.

DOI: 10.21276/IJIPSR.2018.06.03.659

Because sequencing error in next-generation sequencing can be high even after quality control filtering and may confound the disease-marker association, we incorporate the ‘‘blocked pooling’’ design suggested by Wang *et al.* [44] Using this approach, each pooled sample is barcoded, and multiple indexed DNA pools will be sequenced in one lane. This blocked design allows accurate estimation of both locus-specific sequencing error rate and allele frequency.

3) **Association test:** After the individual genotypes are imputed from pooled sequence data, standard logistic regression can be used to test for disease-marker association, with the covariates being adjusted in the model. To take into account the uncertainty of imputed data and obtain valid estimates, multiple imputation techniques will be applied. Specifically, we will repeat the genotype imputation and association test multiple times (typically 5–10), and combine the results to produce estimated effects and confidence intervals [27].

Earp *et al.,* 2011 by examining the variation in allele frequency estimation on SNP arrays between and within DNA pools they determine how array variance and pool-construction variance contribute to the total variance of allele frequency estimation [13]. Their analysis is based on 27 DNA pools ranging in size from 74 to 446 individual samples, genotyped on a collective total of 128 Illumina bead arrays: 24 1M-Single, 32 1M-Duo, and 72 660-Quad. For all three Illumina SNP array types, their estimates of var (array) were similar, between 3-4 × 10-4 for normalized data. Var (construction) accounted for between 20-40% of pooling variance across 27 pools in normalized data. From this, they conclude that relative to var (array), var (construction) is of less importance in reducing the variance in allele frequency estimation from DNA pools.

Future Prospects of Pooled DNA Technology for Crop Improvement

Pooled-DNA sequencing has radically altered the scope of genetics by providing a landscape of ordered genes and their epigenetic states, access to an enormous range of genetic variation, and the potential to measure gene expression directly with high precision and accuracy. Given a suitable cyber-infrastructure, the integration of biological knowledge and models of networks across species, in a two-way flow from crops to experimental species and back again, will begin to generate new layers of knowledge that can be used for crop improvement. One layer is provided by ENCODE-level analyses; although yet to start in plants, these analyses can guide the interpretation of gene function and variation, thus providing new information to inform the prediction of phenotype from genotype. Another information layer is provided by the systems-level integration of gene function into networks, such as those

DOI: 10.21276/IJIPSR.2018.06.03.659

controlling flowering time in response to day-length and over-wintering. These networks have been identified in *Arabidopsis* and rice, with allelic variation in key 'hubs' strongly influencing network outputs. Evolutionary processes, such as gene duplication, and the possible footprints of domestication can be mapped to networks such as those controlling flowering time. Such 'systems breeding' approaches can use the diverse genomic information to increase the precision of which phenotype can be predicted from genotype, thereby accelerating crop improvement and helping to address food security.

The widespread application of genomics and pooled DNA to plant genetic resources will require the generation and analysis of vast amounts of sequence data. These requires new strategies for data generation, storage, analysis and sharing between laboratories [19]. Coordinated efforts to sequence the genetic resources of all major food crops are required. Ongoing advances in DNA sequencing technology can be expected to greatly reduce the cost of genome analysis and make widespread application of genomics to plant genetic resources feasible. DivSeek (*DivSeek.org*) is a coordinated effort to capture global plant diversity for agriculture by genotypic and phenotypic characterization of *ex situ* genetic resources. This initiative would enable crop breeders to access the genomes of all resources in gene banks worldwide and provide an important contribution to global food security.

Conservation of CWRs *in situ* [20] can also be guided by analysis of the genomes of wild populations. Novel wild populations may be identified and prioritized for increased *in situ* conservation efforts and for collection to complement existing *ex-situ* collections. As we move closer to have a reference genome sequence for all of the major crop plant species [31], we need to consider priorities for sequencing of CWR. A reference genome sequence for each of the 1667 priority CWR species [43] would be agood initial target for coordinated international genomics efforts. Genomics combines the tools to work with both current crop species and new species. New species that may be CWR or totally new options may be domesticated with the support of genome analysis [19] & [37] adding to the diversity of crop plants. Analysis of the genome allows targeting of domestication loci in the selection of genotypes for domestication to accelerate the production of a new crop [33]. This approach uses the tools of genome analysis to provide novel options to support global food security that may not be possible without genomics and sequencing technology [33].

CONCLUSION

Global climate change is predicted to impose a severe threat to agricultural productivity worldwide, and thereby challenge food security and nutritional security. Technological advances, particularly transgene-based and molecular breeding technologies have facilitated

DOI: 10.21276/IJIPSR.2018.06.03.659

the development of elite genotypes with durable adaptation to climate change. Pooled DNA-assisted breeding, in particular, is predicted to playing a significant role in the development of climate change resilient crops. Excellent model organisms for climate change have been identified for deciphering traits that need to be decoded and introgressed in the crop plants.

Advances in DNA sequencing technologies and the sequencing of CWR, along with advanced genomics tools will expedite the identification of novel genes and key regulatory regions of stress tolerance toward the development of new cultivars with durable resistance. Although the impact of climate change on crop's resistance is difficult to predict and is likely to be variable depending on the crop and environment, pooled DNA-assisted breeding could contribute significantly to reduce the impact of climate change on future cropping scenarios.

RECOMMENDATIONS

- Establishment of formal interdisciplinary training courses, mainly at the undergraduate level, covering basic scientific molecular techniques such as genomics and sequencing.
- Institution of new programs, at Master's or early postgraduate level with combined crop breeding and scientific training, to support the rapidly developing field of plant biotechnology
- Encouragement of more projects and programs, with some provision for crop improvement, to provide core scientific training for both scientists and breeders in the longer term training courses.
- Plant biotechnology is an area that would benefit from coordination at all levels. Thus, close cooperation between industry, research centers, academia, crop varieties, regulatory bodies, funding agencies, breeders, organizations, investors and other stakeholders could dramatically boost this promising field.

REFERENCES

1. Amos CI, Frazier ML, Wang W. (2000). DNA pooling in mutation detection with reference to sequence analysis. Am J Hum Genet. 66, 1689-1692.
2. Arnheim N, Strange C, Erlich H. (1985). Use of pooled DNA samples to detect linkage disequilibrium of polymorphic restriction fragments and human disease: studies of the HLA class II loci. 82, 6970-6974.
3. Bansal A, van den Boom D, Kammerer S, Honisch C, Adam G, Cantor CR, Kleyn P, Braun A. (2002). Association testing by DNA pooling: an effective initial screen. Proc Natl Acad Sci USA. 99, 16871-16874.

DOI: 10.21276/IJIPSR.2018.06.03.659

4. Barcellos, L. F. *et al.* Association mapping of disease loci, by use of a pooled DNA genomic screen. *Am. J. Hum. Genet.* **61**, 737–747 (1997).

5. Bodmer W, Bonilla C. (2008). Common and rare variants in multifactorial susceptibility to common diseases. Nat Genet. 40, 695-701.

6. Braun, A., Little, D. P. & Koster, H. Detecting *CFTR* gene mutations by using primer oligo base extension and mass spectrometry. *Clin. Chem.* **43**, 1151–1158 (1997).

7. Brozynask, R., Bell, J. Analysis of Multistage pooling studies of biological specimens for estimating disease incidence and prevalence. *Biometrics* **55**, 608–612 (2014).

8. Charles Worth, B. (2010). "Molecular population genomics: A short history".*Genetics Research* **92** (5–6): 397. doi:10.1017/S0016672310000522.

9. Chi A. *et al* (2009). A Two-stage genome-wide association study of sporadic amyotrophic lateral sclerosis. Hum Mol Genet. 18, 1524-1532.

10. Craig DW, Pearson JV, Szelinger S, Sekar A, Redman M, Corneveaux JJ, Pawlowski TL, Laub T, Nunn G, Stephan DA, Homer N, Huentelman MJ. (2008). Identification of genetic variants using barcoded multiplexed sequencing. Nat Methods. 5, 887-893.

11. Darwin, C. R. (1859). The Origin of Species. London: John Murray.

12. Darwin, C. R. (1868). The Variation of Animals and Plants under Domestication. 2 Vols. London: John Murray.

13. Earp, J. G. *et al.* Strategies for mutation analysis of the large multi-exon *ATM* gene using high-density oligonucleotide arrays. *Genome Res.* **8**, 1245–1258 (1998).

14. Edward and Henery (2011). DNA Sudoku--harnessing high-throughput sequencing for multiplexed specimen analysis. Genome Res. 19, 1243-1253.

15. Erlich , L. A. & Smirnov, I. P. Single-nucleotide polymorphism identification assays using a thermostable DNA polymerase and delayed extraction MALDI–TOF mass spectrometry. *Genome Res.* **7**, 378–388 (1997).

16. Fan, J. B. *et al.* Parallel genotyping of human SNPs using generic high-density oligonucleotide tag arrays. *Genome Res.* **10**, 853–860 (2000).

17. Fawcett, R. G., Dolligenger, P. S., Walsh, P. S. & Griffith, R. Simultaneous amplification and detection of specific DNA sequences. *Biotechnology* **10**, 413–417 (2014).

18. Hebert, P. D., Cywinska, A., & Ball, S. L. (2003). Biological identifications through DNA barcodes. *Proceedings of the Royal Society of London B: Biological Sciences*, *270*(1512), 313-321.

DOI: 10.21276/IJIPSR.2018.06.03.659

19. Henry, W. Design and efficiency of selection experiments for estimating genetic parameters. *Biometrics* **27**, 293–311 (2014a).

20. Henry & Nevo LA, J. H. (2014). A Catalog of Published Genome-Wide Association Studies. National Human Genome Research Institute.

21. Iyengar SK. *et al* (2014). Dissection of genomewide-scan data in extended families reveals a major locus and oligogenic susceptibility for age-related macular degeneration. Am J Hum Genet. 74, 20-39.

22. Jackson PL, Slatkin M. (1972). Accounting for bias from sequencing error in population genetic estimates. Mol Biol Evol. 25, 199-206.

23. Kress, W. J., Wurdack, K. J., Zimmer, E. A., Weigt, L. A., & Janzen, D. H. (2005). Use of DNA barcodes to identify flowering plants. *Proceedings of the National Academy of Sciences of the United States of America*, *102*(23), 8369-8374.

24. Krishnan K, Thomson J. (2014). A more powerful test for comparing two Poisson means. J Stat Plan Inference. 119, 23-35.

25. Lachance, G.; England, P. R.; Tallmon, D.; Jordan S.; Taberlet P. (2013). "The Power and Promise of Population Genomics: From Genotyping to Genome Typing."*Nature Reviews* (4): 981-994.

26. Lewontin, R. C. (1974). The Genetic Basis of Evolutionary Change. New York, NY: Columbia University Press.

27. Little, RJA. And Rubin, DB. Allelic discrimination using fluorogenmeic probes and the 5′nuclease assay. *Genet. Anal.* **14**, 143–149 (2002).

28. Liu, Y., Koyutürk, M., Maxwell, S., Xiang, M., Veigl, M., Cooper, R. S., ... & Zhu, X. (2014). Discovery of common sequences absent in the human reference genome using pooled samples from next generation sequencing. *BMC Genomics*, *15*(1), 685.

29. Luikart N, Williams NM, O'Donovan MC, Owen MJ. (2003). DNA pooling as a tool for large-scale association studies in complex traits. Ann Med. 36, 146–152.

30. Manolio TA. *et al* (2009). Finding the missing heritability of complex diseases. Nature. 461, 747-753.

31. Michael & VanBuren, R. W., Paran, I. & Kesseli, R. V. Identification of markers linked to disease-resistance genes by bulked segregant analysis: a rapid method to detect markers in specific genomic regions by using segregating populations. *Proc. Natl Acad. Sci. USA* **88**, 9828–9832 (2015).

32. Michelmore RW, Paran I, Kesseli RV. (1991). Identification of markers linked to disease-resistance genes by bulked segregant analysis: a rapid method to detect

DOI: 10.21276/IJIPSR.2018.06.03.659

markers in specific genomic regions by using segregating populations. Proc Natl Acad Sci U S A. 88, 9828-9832.

33. Malory, T. E. (2011). The new look of behavioral genetics in developmental psychopathology: gene-environment interplay in antisocial behaviors.*Psychological Bulletin*, *131*(4), 533.

34. Nock S, Walker N, Riches D, Egholm M, Todd JA. (2011). Rare variants of IFIH1, a gene implicated in antiviral responses, protect against type 1 diabetes. Science. 324, 387- 389.

35. Nordfors, l. *et al.* Large-scale genotyping of single nucleotide polymorphisms by pyrosequencing and validation against the 5′□ nuclease (TaqMan) assay. *Hum. Mutat.* **19**, 395–401 (2000).

36. Rivas, N. J. Searching for genetic determinants in the new millennium. *Nature* **405**, 847–856 (2011).

37. Shapter SH, Carrasquillo MM, Kashuk C, Puffenberger EG, Chakravarti A. (2013). Allele frequency distributions in pooled DNA samples: appliedtions to mapping complex disease genes. Genome Res. 8, 111-123.

38. Schilling, M. P.; Wolf, P. G.; Duffy, A. M.; Rai, H. S.; Rowe, C. A.; Richardson, B. A.; Mock, K. E. (2014). "Genotyping-by-Sequencing for Populus Population Genomics:

39. Sheffield VC, Carmi R, Kwitek-Black A, Rokhlina T, Nishimura D, Duyk GM, Elbedour K, Sunden SL, Stone EM. (1994). Identification of a Bardet–Biedl syndrome locus on chromosome 3 and evaluation of an efficient approach to homozygosity mapping. Hum. Mol. Genet. Hum. Mol. Genet. 3, 1331-1335.

40. [40] Syvanen, A. C., Aalto-Setala, K., Kontula, K. & Soderlund, H. A primer-guided nucleotide incorporation assay in the genotyping of apolipoprotein E. *Genomics* **8**, 684–692 (1990).

41. Uhl, G., Liu, Q.-R., Walther, W., Hess, J. & Naiman, D. Polysubstance abuse — vulnerability genes: genome scans for theassociation, using 1,004 subjects and 1,494 single nucleotide polymorphisms. *Am. J. Hum. Genet.* **69**, 1290–1300 (2001).

42. Vaughan, P. & McCarthy, T. V. A novel process for mutation detection using uracil DNA-glycosylase. *Nucleic Acids Res.* **26**, 810–815 (1998).

43. Vincent, J. C., Adams, M. D., Myers, E. W., Li, P. W., Mural, R. J., Sutton, G. G., ... & Gocayne, J. D. (2013). The sequence of the human genome. *Science*, *291*(5507), 1304-1351.

DOI: 10.21276/IJIPSR.2018.06.03.659

44. Wang T, Lin C, Zhang Y, Wen R, Ye K (2012) Design and Statistical Analysis of Pooled Next Generation Sequencing for Rare Variants. Journal of Probability and Statistics 2012: 19.

45. Wasson, J., Skolnick, G., Love-Gregory, L. & Permutt, M. A. Assessing allele frequencies of single nucleotide polymorphisms in DNA pools by pyrosequencing technology. *Biotechniques* **32**, 1144–1152 (2002).