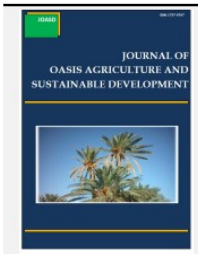

JOURNAL OF OASIS AGRICULTURE AND SUSTAINABLE DEVELOPMENT

www.joasdjournal.org

JOASD

p-ISSN: 2724 – 699X
e-ISSN: 2724-7007

Management and Use of Genetic Resources for Climate Smart Crop Improvement: Overview of the Tailor-Made Training (TMT) Course

Sarrah Ben M'Barek[1,2,3*], Mokhtar El Bekkay[4], Naceur Djebali[5], Chris Kik[6], Alejandro Therese Navarro[6], Angela Machacilla[6] & Daniel Danial[6]

[1]*Laboratory of Support for the Sustainability of Agricultural Production Systems in the North West Region, Higher School of Agriculture of Kef (ESAK), University of Jendouba, Le Kef, Tunisia.*
[2]*Regional Field Crops Research Center of Beja, Tunisia (CRRGC), Tunisia.*
[3]*Wheat Septoria Phenotyping Platform, Tunisia.*
[4]*Laboratory of dry land farming and oasis cropping- Arid Regions Institute (IRA) Medenine, Tunisia.*
[5]*Laboratory of BioActives Substances - Center of Biotechnology of Borj Cédria (CBBC), Tunisia.*
[6]*Wageningen University & Research, (WUR), the Netherlands.*

Article info

Article history:
Received: 03 July 2024
Accepted: 23 July 2024

Keywords: Plant Genetic Resources, Management, policies, Sequencing, Bioinformatics.

***Corresponding author**
sarrah.bm2712@gmail.com

Conflict of Interest: The authors declare no conflict of interest.

Abstract

Via a close cooperation between Dutch and Tunisian research organizations and leveraging the possibilities provided by the Dutch Orange Knowledge Fund, a tailor-made training course on management and use of genetic resources for climate smart crop improvement took place in Tunisia from April 15-26 2024. The course consisted of two modules: one focused on Plant Genetic Resources Management and the other on Sequencing and Bioinformatics. The module on PGR management covered key topics including food security and Biodiversity, PGR management, and PGR policies, whereas the module on Sequencing and Bioinformatics provided up to date information on DNA sequencing methods, genome assembly, genotyping, diversity assessment and genome-wide association studies. The course was structured with a combination of lectures and practical assignments. Additionally, two full-day field visits were integrated into the program, providing participants with valuable information and practical knowledge in on farm management techniques. In total, the course was attended by 26 participants.

1. INTRODUCTION

Tunisia faces significant challenges in ensuring food security amidst the growing impacts of climate change on agriculture, which underscores the urgency of identifying climate-resilient crop varieties. In this regard, the identification and utilization of plant genetic resources (PGR) can provide genetic diversity and enhance crop resilience by incorporating valuable traits such as drought tolerance, disease resistance, and nutritional quality. This, in turn, will improve food security in Tunisia. Hence, effective conservation, management, and utilization of these PGR using modern plant breeding technologies are crucial to addressing this challenge. In this context, Wageningen University & Research (WUR)-Plant Breeding (the Netherlands) in collaboration with the Institute of Agricultural Research and Higher Education (IRESA) have organized and implemented a tailor-made training (TMT) course on Management and use of genetic resources for climate-smart crop improvement. The training course is a continuation of the collaborative efforts between WUR-Plant Breeding and IRESA to further strengthen capacity development in developing new tools and upgrade skills among participating organizations in the management and use of genetic resources in modern plant breeding. The TMT is funded by the Dutch ministry of foreign affairs and is managed by the Orange Knowledge Program (OKP) of NUFFIC, the Netherlands, with the objectives to: ***i.*** upgrade skills on the standard protocols for routine gene bank operations, and ***ii.*** familiarize the participants

with practical procedures for utilizing genetic resources through smart and efficient pre-breeding & plant breeding technologies.
This paper documents a description of the activities covered during the training course.

2. MODULES ADDRESSED IN THE TMT

2.1. Module 1: Plant genetic resources management

2.1.1. The importance of food security and biodiversity

Food security

Two factors significantly influence food security, namely the increasing world population and the effects of climate change on food production. Currently, out of the 7.8 billion people on Earth, approximately 0.8 billion suffer from hunger, 0.5 billion from obesity, and 2 billion from malnutrition. Surprisingly, twice as much food is produced worldwide as is needed in the short term. An in-depth analysis by Smil (2000) revealed that food wastage and meat and dairy production play substantial roles in determining overall food production.Food wastage alone can account for 30% of the post-harvest losses, while meat and dairy production accounts for 26% of these losses. Conversely, food production can be increased by reducing crop losses (estimated at 20-30% annually) and enhancing crop production through improved cultivation and breeding techniques exemplified by the success of the Green Revolution in the 1950s and 1960s. During this period, substantial international investments in agriculture led to significant increases in production of major grain crops such as wheat, barley, oats, and rice (Godfray et al., 2010).
Food security is not solely determined by food availability (e.g. production and distribution of food, landownership, breeding), but also by other critical factors including access (e.g. affordability and distribution of food), utilization (e.g. preparation, processing, cooking and nutritional value) and stability (e.g. consistency in accessing food over time). In a broader context, within food systems, food security is influenced by activities across the food supply chain (from farm to fork), socio-economic factors (e.g. income, employment, social and political capital) and environmental drivers (e.g. climate change, biodiversity, water availability).

Biodiversity

According to Wilson (1988), it was W.G. Rosen in 1985 who first coined the term biodiversity as a contraction of biological diversity. Wilson (1988) indicated that biodiversity can be analysed at different levels, from the genetic level to ecosystems and indigenous knowledge. Furthermore, he explained that biodiversity matters due to its utilitarian values, provision of ecosystem services, and ethical and aesthetic values.
Biodiversity differs from genetic resources, as genetic resources consist of any genetic material of actual and potential value, whereas biodiversity comprises all variation (FAO, 2011). Crop diversity as part of biodiversity, is not randomly spread over the world. The first to discover this was the Russian researcher N.I. Vavilov in the early 1900s, who called these specific regions “centres of origin”. Later, this term was refined to “centres of biodiversity”. Vavilov was also among the first to recognize the potential of landraces (and wild material) for the improvement of existing varieties, which led to the establishment of one of the first gene banks worldwide in St. Petersburg. Nowadays, 7.2 million accessions are held by more than 1,500 organizations worldwide, accessible via GENESYS (www.genesysy-pgr.org).
Biodiversity has been threatened since the last century by habitat change, climate change, invasive species, over-exploitation and pollution, with these drivers affecting different biomes differently (Millenium assessment, 2005). The overall negative impact of these factors suggests that a sixth mass extinction event is currently taking place.
Domestication took place around 12.000 years ago during the Neolithic revolution when a gradual transition took place of a hunter/gatherer society to an agricultural society. This transformation first occurred in the Fertile Crescent, where crops like rye, barley, wheat, and peas were initially domesticated. This combination of crops was not without a reason as they account for an elementary diet. Crops like rye, barley and wheat were selected as carbohydrate sources and crops like peas as a protein source. Some two millennia later this conversion also occurred in Meso-America and China. This domestication process reduced genetic variation compared to their ancestral species, a trend exacerbated by modern breeding practices since around 1900 (Tanksley & McCouch, 1997), leading to genetic erosion.
Crop uniformity can lead to vulnerability. An example of this is the Irish potato famine which took place in between 1840-1850. Due to the use of only a few genetically similar potato clones by

Irish farmers, which proved to be susceptible to *Phytophthora infestans*, this oomycete destroyed the only important food source of the Irish farmers, resulting in the deaths of one million people and the emigration of two million more. Also currently crop uniformity can lead to problems as seen in cases such as banana & Panama and wheat & rust diseases. Even more worrying in this respect, is the fact that the human diet currently depends for around 60% of their calorie intake on just three species—rice, maize, and wheat (FAO, 2011). This is explained by the intense competition in the food market and the drive for cost-effective, large-scale food production within the agro-industrial complex. Uniformity is at a premium during production requiring a few crops and uniform cultivars per crop for efficient food production. Currently,the downsides of this production system are increasingly evident, prompting a search for more sustainable food production methods.

An assignment was carried out to explore the possibilities of GENESYS in obtaining an overview of genetic resources of a species present in Tunisia and elsewhere, and secondly, to establish a collection of drought-tolerant accessions for breeding an improved barley variety (Fig. 1). Furthermore a video entitled Seed Battles (https://www.youtube.com/watch?v=lqarjuqG-XY) was shown and discussed.

2.1.2. Management of genetic resources: the *ex situ* approach and its relation with *in situ* and on farm management

***Ex situ* management**

Ex situ management involves the conservation and use of genetic resources outside their natural habitat. Botanical gardens and genebanks collectively harbour a total of 13.3 million accessions (genebanks 7.2 million accessions and botanical gardens 6.1 million accessions). The major difference between these organizations lies in how the genetic resource exchange occurs. Botanical gardens primarily exchange resources among themselves whereas genebanks engage in exchanges predominantly with both commercial and non-commercial organizations.

Genebank management consists of several processes, namely acquisition, regeneration, characterization and evaluation, seed storage, data management and seed distribution and information exchange (Crop GeneBank Knowledge Base). The acquisition of genetic resources via collecting missions is an important aspect of genebank management as it is the only way to acquire unique germplasm. It was shown that for self- and cross-fertilizing species, the optimal method involves sampling seeds per individual from a population and maintaining them separately (Breese, 1989). Regeneration should similarly occur on an individual plant basis. For self-fertilizing species, this involves selfing individual plants to preserve 100% of the genetic variation present in the original sample. In cross-fertilizing species, a biparental crossing scheme is used to retain most of the variation present in the original sample. Despite these optimal methods for maintaining genetic

Fig. 1. Practical session of use of GENESYS to obtain an overview of plant genetic resources in Tunisia.

variation, genebanks mostly use bulk collecting and regeneration of samples. Depending on the genetic variation in self-fertilizing populations, up to 50% of the genetic variation will be lost after two bulk regeneration cycles (Cross & Wallace, 1994). For cross-fertilizing species, the loss of genetic variation using bulking depends on factors such as allele frequency and the Ne/N ratio. A low frequency of the allele (u=0.05) and a low Ne/N ratio (0.2) leads to the loss of 50% of the allele in three regeneration cycles (Gale & Lawrence, 1984). To reduce the loss of genetic variation during storage in a genebank, one of the options is to limit the number of regeneration cycles. For orthodox seeds, this involves properly drying seeds and storing them in vacuum-sealed bags at -18 °C. Storage of seeds at + 4 °C conditions often lead to significant germination losses, even for species like cereals (van Treuren et al., 2018). It is essential to note that genebanks are not solely repository for conserving seeds but also for distributing them. Therefore, it is important that accessions and

their related information present in a genebank are (online) available and if possible, can be obtained via the standard material transfer agreement (SMTA) of the International Treaty on Plant Genetic Resources for Food and Agriculture (ITPGRFA).

On farm management

Before the turn of the century on farm management was considered as an integral part of *in situ* management. Since then, it has become increasingly evident that on farm management has its own distinctive properties. Brush (2005) defines on farm management as follows: on farm management is the management of domesticated or cultivated species, in the surroundings where they have developed their distinctive properties. Lowder and co-workers (2015) estimated the scale of on farm management worldwide and showed that there are in total around 570 million farms worldwide, with the majority being small (84%), and family-operated farms. Globally small-scale farms (< 2 ha) operate on ca. 12% of the agricultural land, but in low- and middle-income countries, these farms account for 70-80% of total farms operating on 30-40% of the land.

On farm management operates within an informal seed system, whereas conventional farm management takes place in a formal seed system. The primary distinction between these systems lies in the external inputs (e.g. modern varieties, chemical fertilizers, herbicides, fungicides, and pesticides) which are significantly more prevalent in the formal seed system than in the informal seed system. To support and facilitate on farm management community, biodiversity management (CBM) was developed (de Boef et al., 2013). CBM is a community based participatory methodology aimed at strengthening the community's capacity through the management of their knowledge-based systems. The major aim of CBM is to develop ownership and support on farm management of agrobiodiversity and sustainable livelihood options with a minimum of external inputs. Various tools (e.g. four cells approach, social seed network analysis) and practices (e.g. community seed banks, participatory plant breeding) are available to achieve this objective.

***In situ* management**

In situ management involves the conservation of ecosystems and natural habitats and the maintenance and recovery of viable populations of species in their natural surroundings (Heywood & Dulloo, 2005). While the conservation of natural resources has a long history, global biodiversity conservation is a more recent development. Regulation of the use of natural resources (resource ethics) became necessary to prevent selfish motives and thus overexploitation that could compromise long-term sustainability, a social dilemma often called the Tragedy of the Commons. The origin of modern *in situ* management finds its roots in the Enlightenment period in Europe (late 18th century). Conservation ethics were formulated during that period and practically applied to forests in British India (1855) and later in the USA (Yellowstone Park, 1872). The term conservation came into widespread use in the late 19th century and referred to the management, mainly for economic reasons, of timber, fish, game, and secondly for the preservation of forests, wildlife, and wilderness. The target of conservation in the 1950's was mostly on protecting individual species (by protecting the panda, you also protect its habitat). Nowadays the ecosystem approach has taken-over the single species and single habitat approach. Since the 1950's, important international meetings took place to anchor *in situ* management into society: 1972 - UN: adoption of a program to promote conservation of unique sites to the common heritage of mankind, 1978 – First international conference of Conservation Biology, 1992 – Adoption of the Convention of Biodiversity (CBD). Major components in the development of a conservation strategy for a region were formulated during these conferences.

The complementarity of the various management methods

In situ and on farm management facilitate evolutionary processes, whereas this is less the case in *ex situ* management. Furthermore, the sovereignty of the seeds is much better guaranteed via *in situ* and on farm management compared to *ex situ* management. On the other hand, security, accessibility, and characterization & evaluation are better guaranteed via *ex situ* management then via *in situ* and on farm management. Therefore, these three genetic resources strategies can be seen as complementary and the use of all three strategies is encouraged in the management of genetic resources.

An assignment was carried out to assess the role of the three genetic resources management

strategies in international efforts to promote coconut genetic resources conservation and use. In this context the summary of the global coconut strategy (COGENT, 2018) was used.

2.1.3. Plant genetic resources (PGR) policies

2.1.3.1. Intellectual Property Rights (IPR)

Plant Breeders Rights (PBR) and patents

On the interface between a commercial product with value and a technology, Intellectual Property Rights (PBR and patents) and their international agreements (UPOV and TRIPS) are present.

Plant Breeders Rights (PBR) are the rights granted to the breeder of a new plant variety giving the breeder exclusive control over the propagating material of a new variety for 20-25 years. To qualify for PBR, a variety must be distinct, uniform, and stable, often referred as the DUS criteria. Common exemptions of PBR include the ability for breeders to use protected varieties to develop new varieties (breeders' exemption) and farmers to use farm-saved seed of protected varieties (farmers' privilege). Before being allowed to the market, a candidate variety must undergo a certification process, which typically takes several years.

A patent is a set of exclusive rights granted by a sovereign state to an inventor for a limited period in exchange for the public disclosure of the invention. Claims must meet novelty, non-obviousness and have and industrial application. Patents, like PBR, are also granted for 20-25 years. Patents can give the owner the right to prevent others from using the patented invention without permission. There is a difference between patents granted in the US and EU as US granted patents allow for the patenting of any living organism that is a product of human intervention, whereas in case of patents granted in the EU, plant or animal varieties or essential biological processes can be patented for their production but are very hard to obtain due to severe juridical thresholds. Patents costs money as for every country where the patent is active, annual fees must be paid. Costs will rise significantly if the patent is challenged by another party. Reasons for universities or research groups to apply for a patent may include agreements with subsidizing parties or the opportunity to sell the patent or licence to companies in return for funding. However, in most cases, there is no desire to maintain patents due to their hight maintenance costs. In the US, a variety can be protected by PBR and patents, whereas this is not possible in the EU.

International IPR agreements

There are currently two international IPR agreements active, namely the UPOV convention and the TRIPS agreement.

UPOV (Union Internationale pour la Protection des Obtentions Vegetables) aims to create an international regime to protect IPR over plant varieties by means of PBR. It was set-up by European breeders to keep breeding out of the US patent system. UPOV came into force in 1961 and has been revised several times, leading to increasingly weaker protection of farmers' rights. Under UPOV 1991, farmers are not allowed to freely exchange seeds of protected varieties and national authorities may decide that farmers no longer can use saved seeds or require payment of royalties. There are currently 74 countries that are UPOV members.

TRIPS (Trade-related Aspects of Intellectual Property Rights) aims to establish a minimum set of international standards for the protection of IPR. It was negotiated in the framework of the World Trade Organization (WTO) and entered into force in 1996. There are currently 164 parties to the TRIPS. TRIPS stipulates that patents that are granted to inventions, whether products or processes, in all fields of technology, as long as they are new, involve a creative step, and have an industrial application. However, members may exclude plants and animals from patentability if they provide another efficient protection system. Often countries have chosen in this case to follow the 'ready made' UPOV model to protect plant varieties.

2.1.3.2. Access and Benefit Sharing

On the interface between a commercial product with value and genetic resources, ABS and its international agreements are present. Access and Benefit Sharing (ABS) regulate the access to (and use of) genetic resources and associated information, and the sharing of benefits from this use between providers and users. Important aspects involve benefit sharing and farmers' rights. Possible benefits include monetary (royalties, fees, funding of research) and non-monetary (capacity building, scientific cooperation, technology transfer) aspects. Farmers' rights usually mean firstly to include the right to save, use, exchange and sell farm-saved seeds. Additionally, they involve recognition and reward for their contribution to the global pool of genetic resources as well as

the development of commercial varieties. Farmer's rights also include the right to participate in decision making on issues related to genetic resources.

International ABS agreements

Currently there are two international arrangements on ABS, namely CBD/Nagoya and the International Treaty.

The Convention on Biological Diversity (CBD) was negotiated in the United Nations Environment Program (UNEP). Its objectives are conservation of biological diversity, sustainable use of its components and the fair and equitable sharing of benefits arising from the utilization of genetic resources. The CBS entered into force in 1993. Currently there are 196 parties to the CBD. An important element of the CBD is the concept of national sovereignty, which implies that countries can require that collectors must obtain permission before access to genetic resources. The ABS takes place through bilateral contacts and on a case-by-case basis. Access must be granted on mutually agreed terms (MAT) and prior informed consent (PIC). The CBD is primarily focussed on regulating access to genetic resources of wild flora and fauna used for chemical and pharmaceutical purposes as chemical and pharma companies made large profits without substantial benefit sharing. The Nagoya protocol was developed to further advance the implementation of the third objective of the CBD and entered into force in 2014. The Nagoya protocol contains provisions aiming at more predictable and transparent conditions for access to genetic resources. It also contains provisions on access to traditional knowledge, compliance with domestic ABS regulations and MAT and awareness raising and capacity building. The Nagoya protocol does not apply when other international ABS instruments apply such as the Treaty.

The International Treaty on Plant Genetic Resources for Food and Agriculture (ITPGRFA) has been negotiated within the Food and Agriculture Organisation (FAO). It is the result of the renegotiation of an earlier international agreement, namely the International Undertaking, which was adopted in 1983, and bringing it in harmony with the CBD. The objectives of TRIPS are conservation of biological diversity, sustainable use of its components and the fair and equitable sharing of benefits arising from the utilization of genetic resources. It entered into force in 2004 and has currently 139 parties. Important elements are: ABS goals need to be achieved through a multilateral system (MLS) of exchange of genetic resources, secondly a list of crops has been established which are covered under the MLS, thirdly the Treaty focuses less on chemical and pharma companies but more on the agro sector, regulating the access to land/collections under the control of the government, and fourthly facilitated access is based on a Standard Material Agreement (SMTA). Benefit sharing in the Treaty encompasses: (i) facilitated access which is a major benefit in itself, (ii) benefits arising from the use of PGRFA must be shared through the exchange of information, access to and transfer of technology, and capacity building, and (iii) the sharing of monetary and other benefits arising from commercialization of which part of the monetary benefits must be placed into an International Benefit Sharing Fund, which is meant to support conservation and crop improvement efforts especially by small-scale farmers in developing countries.

2.1.3.3. Digital sequence information

Digital sequence information (DSI) is based in general terms on the sequence information of the genome, which allows researchers to study or use a genetic resource with accessing the physical resource. Until present DSI is not regulated and therefore ABS can be circumvented, which can lead to unwanted situations like in case of *Stevia rebaudiana*.

In 1996, DSI appeared for the first time on the CBD agenda and also later on the Treaty agenda. This led in 2022 to an agreement on the establishment of a multilateral mechanism for the use of DSI on genetic resources. In 2023, an *ad hoc* working group on DSI was established to undertake further development of the multilateral system.

An assignment was carried out with the aim to understand which ABS regulations are implemented in Tunisia and which role the implemented ABS regulations play in organizing a collecting mission by researchers from another country.

2.1.4. Field studies

Two field days were organized. The first that was proposed by Dr. Ramzi Chaabane (NGBT, Tunisia) took place to The Lella Kmar Cooperative located in Jedaida in the governorate of Manouba (Fig. 2). Created in 2019, the association mainly works in agribusiness and the processing of indigenous seed-based products. They have about 330 hectares that are organically certified, including

Fig. 2. Visit to Lella Kmar Cooperative located in Jedaida in the governorate of Manouba.

160 hectares of wheat from a landrace called Mahmoudi from which they produce semolina and couscous. The remaining land is used for olive trees and medicinal and aromatic plants.
The association encourages its members to grow "seeds of the local landraces that are better adapted to the climate and Tunisian soils, thus adding value to food security and nutrition. Their objective is to support farmers in their transition towards organic and indigenous seeds assisting those who own small plots of land by providing seeds and farming equipment to avoid land abandonment. The cooperative has 300 members, 80% of whom are women, aiming to empowerwomen groups and promote economic development in rural areas. The Lella Kmar Cooperative provides employment opportunities and income generation for local women as most of the seed-based products are exported to France with Ethiquable or commercialised at the local market.
In addition, by participating in the cooperative, women can contribute to their households' incomes, gain economic independence, and access training and support to enhance their skills in production, communication, and marketing.
The second field day was proposed by Prof. Messaoud Mars (ISACM, Tunisia) specialized in fruit trees and particularly Figs, and took place in the governorate of Beja, a region called Djebba, where Figs and other fruit trees play a significant role in local agriculture and nutrition (Fig. 3).

Fig. 3. Visit to region of Djebba in the governorate of Béja where local varieties of figs and other fruit trees are maintained (in-situ conservation).

The genetic biodiversity of Figs in Djebba, provides a foundation for cultivating varieties that are well-adapted to the region's climate and soil conditions. By leveraging this diversity, farmers in Djebba can enhance food security by cultivating Fig varieties that are more resistant to pests, diseases, and adverse weather conditions. Additionally, Figs are a nutritious fruit rich in fiber, vitamins, and minerals, contributing to improved nutrition and health outcomes in the community.
Both field days included assignments and were beneficial to the participants to gain more experience and knowledge from lessons learned on how genetic resources are maintained and conserved in *ex situ, in situ* and on farm.

2.2. Module 2: Sequencing and bioinformatics

Sequencing technologies have slowly revolutionized the field of genetics spanning from biomedicine to ecology. Agriculture has not

lagged, with numerous agricultural genetic studies relying on the acquisition and use of sequencing and genotyping information. During this course, students delved into the computational aspects of sequencing technologies including library preparation and sequencing, genome assembly, diversity studies and genome-wide association studies (GWAS). These topics were explored both theoretically and practically during the tailor-training course.

2.2.1. DNA sequencing methods

Since the invention of the Sanger sequencing reaction, many next-generation sequencing technologies have been developed. The Sanger technique was extremely limited in its throughput, meaning that to sequence the first human genome took over a decade and $2.7 billion to complete(Hu et al., 2021). Next-generation sequencing (NGS) approaches aimed at overcoming this limitation using different types of parallel sequencing: simultaneous sequencing of several molecules at once. The increase in output was dramatic, where a single Sanger reaction would provide information about ~1000bp, a modern Illumina sequencing experiment can generate 1.8Gbp. However, this improvement came with a drawback of reduced accuracy. All modern NGS methods have higher error rates than the Sanger reaction, although with the appropriate data analysis the impact of errors can become negligible. An important aspect in that regard is sequencing depth: the number of times a specific DNA fragment is sequenced (Sims et al., 2014). The higher the depth, the more accurate sequencing becomes. Library preparation, the pre-processing of DNA, is therefore crucial since an appropriate library can greatly increase the usefulness and accuracy of a sequencing experiment.

During the first day of the course, we did a quick overview of sequencing technologies and library preparation methods to discuss the types and number of errors that can be expected when using different combinations of library and sequencing technology.

Library preparation

All NGS methods require some form of DNA processing known as library preparation.

The first step is fragmentation. No NGS method can sequence the whole genome at once, thus shorter fragments of homogeneous length should be obtained. Fragmentation can be performed using several methods, but the most common are physical fragmentation through sonication, and enzymatic fragmentation using specific or unspecific restriction enzymes. The process of fragmentation can be modified to reduce the number of genomic regions that contribute to the final DNA sample. Whether this reduction is targeted or untargeted is the main difference between two popular approaches: reduced representation libraries and targeted sequencing.

Reduced representation libraries (RRL) use restriction enzymes to create DNA fragments with specific sequences at each end of the DNA fragment. This step is followed by a specific amplification for fragments with a specific restriction site combination, thus vastly increasing the concentration of a small subset of random DNA fragments throughout the genome. Several variations of this protocol have been developed over the years, although two of the most popular are genotyping-by-sequencing and RADseq (Chung et al., 2017; Elshire et al., 2011; Miller et al., 2007).

Targeted approaches, on the other hand, offer the possibility of choosing which pieces of DNA are selected. Again, several methodologies have been developed for this purpose, but the most common ones rely on complementary DNA probes that hybridize with specific fragments, which are then separated from the total genomic DNA (Cronn et al., 2012; Hu et al., 2021). The challenge here resides in the probe design step since it requires substantial genomic information from the organism under study.

2.2.2. Sequencing technologies

An important aspect of sequencing experiments is the approach used. There are four main companies, each with their own proprietary method to sequence DNA. Their cost, sequence length and error rate are the main determinants when deciding which technology to use. These four techniques can be divided in second and third generation sequencing which produce short and long reads respectively (Table 1). For a full review of these techniques, see Goodwin et al. (2016):

- Illumina: the most-used method of short-read sequencing, it produces short reads with a very high throughput and is the most accurate technique within the NGS methods. Their technology is based on solid-phase bridge amplification which enables sequencing of DNA fragments from both ends. As a result, Illumina produces “paired-end” information: pairs of reads that correspond to the left and right sides of a fragment, with accurate estimations of the

Table1. Comparison of sequencing technologies

	Sanger	Illumina	IonTorrent	PacBio	Nanopore
Error	>0.001%	0-1.5%	0.5-1%	10% (pass); >0.001% (CCS)	5.5% – 8%
Length	800-1000bp	150-400bp	200-400bp	10kb – 100kb	500kb – 1Mb
Throughput	1 PCR product	DNA/RNA library	DNA/RNA library	DNA/RNA library	DNA/RNA library
Cost per bp	0.2-1	0.01-0.1	0.1-0.5	1-5	~1$

All measures according to the most recent iterations of each technology. Sources: Illumina (Stoler&Nekrutenko, 2021); IonTorrent (Song et al., 2017); PacBio (PacBio, 2024); Nanopore (Delahaye & Nicolas, 2021).

number of bases between the two fragments. This information is often used during read mapping and assembly.

- Ion-torrent: the principle of this short-read method is to detect the protons released during DNA synthesis using a semiconductor plate at the bottom of a sequencing well. The amount of voltage detected when adding nucleotides determines the number of bases added during sequencing, a process that is prone to errors in homopolymer regions with more than 7 identical bases after each other.
- Single-Molecule Real Time (SMRT): the long-read method developed by Pacific Biosciences, also known as HiFi sequencing, is based on circular DNA sequencing. The two complementary strands are sequenced repeatedly, each time being known as a "pass". The error rate within one pass is quite high (~10%), but by sequencing the same strand multiple times a consensus circular sequence can be obtained, with error rates as low as 0.001%. With this technique a balance must be struck between accuracy and length: the longer the sequence the lower the accuracy.
- Nanopore: this long-read technology is quite striking in its method. It is based on a nanopore protein which can translocate DNA through a membrane. As the DNA is translocated the electrostatic signal of the nucleotides is detected and translated into a sequence. This is done through a decoder program called basecaller. Although translating the electric signals into nucleotides is still quite inaccurate, years of fine-tuning have greatly improved base calling accuracy. Further improvements in the future may mean that the same data can be re-analysed to obtain better sequences.

2.2.3. Genome assembly

One of the main uses of sequencing technologies is to assemble genomes. Genome information often paves the way for biotechnological research in crops, helping assess diversity and study functional genetics. All staple crops and most industrially grown crops of the world have had their genomes assembled. As this technology becomes more accessible, it is now feasible to assemble the genomes of less well-studied organisms. For this reason, we had a short overview of genome assembly, what it entails and what are its main limitations.

Two major assembly algorithms have been developed throughout the years: the overlap-layout-consensus (OLC) and the *de Bruijn* graph. OLC is based on three steps: finding overlap between reads, creating a layout (i.e., an order) of the reads based on the overlap, and finding consensus across overlapping reads to correct sequencing errors. The *de Bruijn* method solves the same problem with a different parametrization: it subdivides the sequences in k-mers, subsets of the original reads containing k letters. Each k-mer is connected to all other k-mers with which it overlaps by k-1 letters, thus creating a graph structure known as the *de Bruijn* graph. Paths through this graph represent assemblies. While both methods produce similar accuracies, particularly when data volumes become very large, the *de Bruijn* approach is much more efficient, thus being able to more easily manage the large volumes of reads obtained with NGS methods (Sohn & Nam, 2016).

Nevertheless, both methods struggle with repeated sequences. When reads are shorter than a repeated DNA segment, the resulting overlap is ambiguous, meaning it is not easy to separate the two (or more) repeated sequences in the genome, often leading to fragmented assemblies or chimeric sequences. The most effective way of overcoming this issue is to generate longer reads. Long-read sequencing was precisely developed for this issue and has become an almost essential tool to achieve good quality in genome assemblies (Giani et al., 2020).

Nowadays, high-quality genomes can even aim to generate telomere-to-telomere genomes, complete chromosomal sequences without gaps or breaks(Li & Durbin, 2024). To achieve this, a range of technologies must be used, the most common approach being a combination of long reads to organize the sequence, and short reads to polish the quality. The recalcitrant regions that produce few breakages are then resolved using other techniques such as the chromosome conformation capture offered by Hi-C (Li & Durbin, 2024).

Regardless of assembly method, genome quality can be assessed using several parameters. The most common are the N50, L50 and BUSCO parameters. N50 and L50 refer to the number of contigs (contiguous sequences) and their size: N50 is the length of the shortest contig in the top half of longest contigs, while L50 is the number of longest contigs that represent half of the total assembled sequence. In practice, one desires a large N50, meaning that most contigs are long and a low L50, meaning that the assembly is composed of few fragments. While N50 and L50 evaluate the contiguity of an assembly, BUSCO evaluates its completeness. The acronym stands for "Benchmarking Unique Single-Copy Orthologs" and represents the percentage of housekeeping, single-copy genes present in the assembly. If the genome is not complete, some number of genes will not be found, or will be fragmented, indicating poor genome quality. Nevertheless, assessing genome completeness is a complex issue and several methods for this aspect have been proposed (Van Bel et al., 2019).

2.2.4. Genotyping

Novel technological developments have enabled a much deeper understanding of genetic diversity. During the course, we covered the design of two common types of genotyping approaches: SNP arrays and sequencing-based genotyping. Their popularity stems from their high-throughput and cost-effectiveness, enabling the survey of thousands or millions of polymorphisms for a wide range of samples in a short amount of time.

SNP arrays can simultaneously genotype dozens of thousands of pre-determined markers with extremely high accuracy. Each SNP is scored by detecting hybridization to a set of probes, which are usually designed after assessing a diversity panel of the target germplasm (Batley et al., 2018). Choice of the diversity panel is crucial to avoid ascertainment bias (Geibelid et al., 2021). Although SNP arrays remain the cheapest and most accurate genotyping platform, their bias remains challenging to overcome, which has led many to favour sequencing-based genotyping instead.

High-throughput genotyping can also be achieved using sequencing data. Most often reads are aligned against a reference genome and polymorphisms are detected by comparing the reads against the genome. The main hurdle with this technique is to distinguish between polymorphisms and sequencing errors. With very low depth; the difference between both becomes difficult to distinguish, thus, often statistical methods are used to estimate the most probable genotypes (Chung et al., 2017; Deschamps et al., 2012).

2.2.5. Diversity assessment

The most valuable resource in plant breeding is often considered to be genetic diversity. Genebanks around the world are tasked with safeguarding it, while international treaties like the Convention on Biological Diversity and the Nagoya Protocol regulate its maintenance, use and just distribution. It is through genetic diversity that new plant varieties resistant to diseases and environmental stresses can be obtained. The first step, however, is to characterize this diversity.

During the course, we broadly discussed a set of statistical techniques often used to characterize diversity. All these techniques fall under the broad field of clustering: grouping of samples according to their similarity. Grouping can be based on phenotypic or genotypic data and although often the results are similar that is not always the case.

Phenotyping data consists of phenotypic scores which we can classify into two broad categories: quantitative and qualitative phenotypes. Quantitative traits are naturally numeric, while qualitative traits represent categories. Qualitative traits can be further classified into ordinal, when they follow a natural order; binary, usually representing absence/presence traits; or nominal, when they are simply unordered categories, such as color. For classification purposes quantitative traits are the most informative, followed by ordinal, then binary and finally nominal traits. On the other hand, genotypic data is often expressed numerically, and can be considered a type of quantitative trait.

Four main techniques that were discussed in the module 2 are as follows:

1. Hierarchical clustering: this popular technique is based on distance matrices; therefore, the choice of an appropriate distance metric can dramatically change the results obtained. Hierarchical clustering is expressed in the form of a dendrogram. These are also the basis of phylogenetic classification, a type of clustering based on the similarity matrices obtained through multiple sequence alignment (Kapli et al., 2020).
2. Principal Component Analysis: although not strictly a clustering technique, PCA is well known due to the very interpretable visualizations it produces. PCA is a dimensionality reduction technique that summarizes correlated information across variables into non-correlated principal components. As such, it can easily create maps of the genetic diversity in two dimensions that highlight the similarities between samples and any obvious groupings.
3. K-means clustering: this unsupervised clustering method is quite popular due to its versatility. By defining k centroids, samples can be divided into k groups of high similarity. Deciding the number of centroids to use is often not obvious, in most cases several values of k will be used and an evaluator such as the within sums of squares will be used to find the optimal number of groups.
4. Admixture models: STRUCTURE is a popular software tool which applies a mixture model to assign individuals to k ancestral populations (Pritchard et al., 2000). Mixture models as that are still used nowadays and have the interesting property of assigning groups probabilistically. Thus, a sample can have 0.4 probability of belonging to group A, and 0.6 of belonging to group B. In a genomic context this is equivalent of having 40% of the genome from ancestral population A and 60% from ancestral population B. This type of models can only be used with genetic data but are commonly used descriptors of genetic diversity.

The different techniques highlighted during this part of the course were already of great use to the participants, since many problems common to plant breeding and the management of genetic resources can be formulated as clustering problems. For instance, a group of students was interested in the classification of date palms according to simple sequence repeat (SSR) markers. With the use of PCA and hierarchical clustering they were able to separate the accessions and find 6 of the SSRs which were completely associated with the grouping.

2.2.6. Genome-Wide Association Studies (GWAS)

Since their first successful appearance through the human genome project, GWAS studies have become a hallmark of crop research (Ikegawa, 2012; Thomas et al., 2005). Through the association between phenotype and genotype in large association panels, this approach has proven extremely informative to find causal variation for observed phenotypic diversity. A crucial aspect of GWAS is to consider the genetic structure produced by linkage disequilibrium patterns in the genome, otherwise false positive associations will be found. The most common method to deal with this issue is the mixed model in which a random term accounts forthe genetic structure (Yu et al., 2006).

There are several modifications of the basic mixed model GWAS that have been proposed to tackle subpopulation structure, multi-environment trials or complex experimental settings. In the course we focused on multivariate traits such as those produced by mass spectrometry or multi-year trials. We also covered the definition and application of multiple testing thresholds, a relevant topic regardless of the model used.

Multiple testing and significance thresholds

Although GWAS are often thought of as single experiments, they are composed of hundreds to thousands, sometimes even millions of statistical tests: each marker is tested independently. In most tests of hypotheses, one would apply the 0.05 threshold, if the p-value of the test is below that threshold the null hypothesis (often equivalent to no association) is discarded. This threshold of 0.05 is equivalent to accepting a 5% error rate, which in a single test is reasonable. However, if 1000 tests are performed, this would be equivalent to 50 falsely significant tests (when no association is present). Thus, to avoid too many erroneous results, the significance threshold is modified. We reviewed two methods of performing this modification:

- Bonferroni threshold: this extremely popular and simple method proposes simply to use as threshold 0.05/n where n is the number of tests performed, or number of markers in a GWAS. It is a quite stringent threshold due to the assumption that each test is totally independent, something that does not occur in GWAS due to linkage between nearby markers. Nevertheless,

it remains one of the most used methods for significance detection in multiple testing scenarios.

- Permutation threshold: this is arguably the most accurate threshold estimator, although it is also the most computationally demanding. Variations of this idea exist, all based on permuting the phenotypes and genotypes and determining the baseline p-value distribution under the null hypothesis (Churchill & Doerge, 1994; Uffelmann et al., 2021). These types of tests also help highlight situations where skewed phenotype distributions are generating artificially inflated p-values.

Multivariate trait studies

Crop experiments are often complex trials involving multiple locations and multiple traits being studied at once. Consequently, each studied sample has several trait measurements for different phenotypes and/or locations. These types of datasets are multivariate, and they can be analysed using several statistical techniques to obtain interesting information such as environmental effects, genotype by environment interaction (GxE) or trait coregulation.

The study of environmental and GxE is often done using variance decomposition approaches, in which part of the variation is explained using a mixed model including a genetic component (G), an environmental component (E) and an interaction component (GxE). BLUP and GBLUP estimates make use of this setting and provide accurate estimators of the G component, enabling heritability measurements and selection of "best genotypes". On the other hand, GxE estimates offer an insight into adaptations to specific environments, something of particular interest when dealing with harsh or stressful environments such as semiarid and arid fields. This topic is quite large and has been covered in multiple review books and articles (Beavis et al., 2023; Beavis & Mahama, 2023; Zhang et al., 2021).

It is often also interesting to study the correlation among related variables. In biological experiments often multiple variables are correlated due to sharing some common causes, which can be illuminating during their analysis. For instance, several traits sharing the same significant genomic region highlight a common genetic mechanism behind all traits. In metabolomic data that is often the case, where one genetic regulator influences the concentration of several metabolites. Correlation matrices are in this sense useful, but more sophisticated approaches such as correlation clustering or factor analysis provide more accurate descriptions of the relationships (Rosato et al., 2018).

2.3. Module 3: Training evaluation and conclusions

2.3.1. Sharing the local Knowledge

Four regional and local experts were invited to share their experiences with the participants. The experiences shared included different crops such as cereals, faba beans and Figs grown in different parts of the country. In addition, the experience of using the FIGS methodology in identifying traits to be used in plant breeding were discussed by ICARDA. Four lectures were presented (Fig. 4), namely i. strategies of enhancing the use of genetic resources by breeding programs by Kehel Zakeria, ICARDA; ii. conservation and management of genetic resources in the National Genebank of Tunisia by Ramzi Ben Chaabane, Tunisia; iii. i*n situ* and *ex situ* conservation and valorization of faba bean (*Vicia faba* L.) landraces by Khalil Khamassi, Tunisia; and iv. conservation and use of genetic resources of underutilized fruit crops, case of Fig tree by Messaoud Mars, Tunisia.

Fig. 4. Presentation of the resource persons sharing their local knowledge and expertise

2.3.2. Outcomes of the TMT

- Curriculum and teaching materials were made available to all participants.
- Updated the knowledge of 26 participants (Fig. 5) concerning on how to handle genetic resources and use them in modern plant breeding technology, especially in sequencing technology and data management.

WORKING GROUPS			
GROUP 1	1	Elbekkay Mokhtar	VEGETABLES
	2	Najla Mezghani	
	3	Soumaya Arraouadi	
	4	Naceur DJEBALI - LEADER	
GROUP 2	1	Cyrine Robbana	CEREALS
	2	Daaloul Olfa	
	3	Rania DJEMAL	
	4	Rim Nefissi Ouertani - LEADER	
	5	Warda Jendoubi	
	6	Salem Marzougui	
	7	Sarrah Ben M`Barek	
GROUP 3	1	Elyes Babay - LEADER	legumes / forage
	2	Rouissi Mustapha	
	3	Houda Chennaoul Kourda	
	4	Rim Hajri	
	5	Amine Slim	
GROUP 4	1	Jellouli Neila	medicinal plants/ornamentals
	2	Rhimi Awatef - LEADER	
	3	Hnia Chograni	
	4	Sana Medimagh	
GROUP 5	1	Ahmed Othmani	Olive/figs / citrus / palm date
	3	Azizi Thouraya	
	4	Kadri Karim	
	5	Selma Ben Adbelaali - LEADER	
	6	Wisal Derouich	
	7	Olfa Saddoud	
	8	Zouhour Ouali	

Fig. 5. List of participants in the course and the working groups according to the breeding crops (Refer to Supplementary File 1 for additional information).

- Based on the input provided during the TMT, five action plans (Fig. 6) are developed, namely ondurum wheat, palm dates, Cucurbitaceae, legumes & forage and medicinal plants.
- Strengthened links between participants and trainers through follow-up and developing internship programs for pre-graduates.
- Certificates were handed over to the participants who completed the training course.

2.3.3. Evaluation and Monitoring of the TMT

Summary of the evaluation made are presented below as follows:

1. On scale 1 to 5 (1.0 =poor and 5.0=excellent), an average of 4.6 out of 5.0 was given for the quality of lectures, contents, quality of presentations, and teaching materials, benefits and lessons learned from assignments, and quality of local knowledge.
2. Deep understanding among participants with management of genetic resources and sequencing technology, genotyping, and phenotyping methodology.
3. Application of new tools and use of software to study genetic diversity and for identifying desired traits.
4. The TMT addressed the practical needs of management of genetic resources.
5. The participants expressed their concern about lack of funds to ensure sustainability of the planned activities.
6. Many docents are committed to use the provided teaching materials at their classroom to disseminate the knowledge gained to students and postgraduates.
7. Dissemination of knowledge gained is expected to take place from the farmer communities to the National Genebank of Tunisia.

2.3.4. Impact on local research

Several ongoing research projects can benefit from the lectures and assignments presented in this course.

For module 1 on PGR management, the use of GENESYS to explore the wealth of accessions present in genebanks worldwide is pivotal for setting up various projects related to collecting and/or breeding. The same yields for PGR management strategies, how to integrate *ex situ*, *in situ* and on farm management in a local context and finally PGR policies, a deep understanding of the current policies is of utmost importance for the exchange of PGR within and between countries.

Fig. 6. Discussion of action plan and reflection of field trips programs.

For module 2 on Sequencing and Bioinformatics, the maintenance of traditional landraces of wheat is an important topic for the National Genebank of Tunisia (NGBT). By genotyping and assessing the genetic diversity of these landraces the NGBT could demonstrate the unique genetic diversity present in these populations, as well as test and classify unknown populations to identify to which of their traditional landraces they belong (or whether they represent a mixture of races). Similarly, projects on date palm pollinators have shown that males from different genetic backgrounds produce higher quality and drought resistant dates and classifying males according to their genetic background could guide the male reproduction

efforts and development of new breeding programs focused on pollen quality.

During the course days, it also became evident that some of the modern technologies described are not readily applicable in Tunisia. The main limitation described by the participants is the lack of access to genotyping and sequencing services, which prevent the usage of the most modern genetic analysis techniques. Moreover, without their own datasets being available, it is often difficult to develop the computational skills required to fully utilize the most modern statistical approaches. The practical sessions highlighted that not many researchers had experience in bioinformatic analyses, although they were highly interested in their results. Development of bioinformatic services and skills seemed an important point for the development of the Tunisian plant genetics research, particularly in a country where ample traditional and underexploited genetic resources are still preserved. In the future, courses that offer more foundational bioinformatic training will likely be very well received and provide an interesting avenue of scientific development for Tunisian plant genetics research.

2.3.5. Recommendation for future activities

- More time should be allocated to practical exercises in R-programming and management of genomic data.
- Public-private partnerships are recommended for efficient seed multiplication and distribution.
- The participants recommended efficient use of the existing genetic resources at the National Genebank of Tunisia.
- There is a need for training on:
  - Plant breeding for biotic and abiotic stress factors
  - Bioinformatics and data management
  - Breeding for quality and nutrition
  - Use of molecular tools in plant breeding
  - Improved technology in phenotyping and genotyping
  - Selection methods technology
  - Improved seed technology and improved interaction with farmer communities
  - More case studies on the use of native materials in plant breeding

2.3.6. Sustainability of the training

- Follow-up meetings with the participants are planned after six months to discuss progress made and make the required follow-up.
- Improved links between the participants for sharing knowledge and information.
- Strengthening the links between WUR trainers and the participants through mutual knowledge exchange of knowledge, information sharing and the development of internships.

2.3.7. Action plan recommendation

Based on experience gained and lessons learned in Durum wheat from Lella Kmar cooperative, we highly recommend the upscaling of the Durum wheat production and involving more farmers to improve their welfare and food security. Over the years, the cooperative has:

- Successfully implemented organic farming practices.
- Developed a robust model for cultivating the landrace durum wheat Mahmoudi.
- Established effective processing techniques for producing high-quality semolina and couscous.
- Built strong partnerships with private companies and research institutions.
- Gained valuable insights into optimizing farming practices and expanding market reach.

Lella Kmar Cooperative currently operates on 330 organically certified hectares, including 160 hectares, dedicated to the landrace of durum wheat variety Mahmoudi for semolina and couscous production.

Our challenge is to scale up the production of the cultivated area by involving up to 1,000 farmers aiming at improving nutrition, food production, and creating more job opportunities which benefit a broad range of stakeholders. This expansion should involve partnerships with private companies for ensuring high quality seed production and efficient distribution of required inputs. In addition, involvement of knowledge institutions and NGO is particularly important to address gaps identified and contribute towards the improvement of the agronomic practices. By increasing production and integrating value-added processes, we expect to generate employment opportunities and foster economic growth within the local farming community.

Acknowledgments

On behalf of the Department of Plant Breeding, Wageningen University and Research (WUR), we would like to thank the following entities and persons:

- The Dutch Ministry of Foreign Affairs and Nuffic team for their financial contribution to this training course.
- The Institution of Agricultural Research and Higher Education (IRESA) (DG Prof. Hichem Ben Salem) for their excellent support and for facilitating the implementation of this TMT.
- The trainers for taking part in the event.
- The resource persons -Prof. Messaoud Mars (The Higher Institute of Agronomic Sciences of Chott Mariem), Dr. Zakaria Kehel (The International Center for Agricultural Research in the Dry Areas), Dr. Ramzi Chaabane (The National Genebank of Tunisia), Dr. Khalil Khamassi (National Institute of Agronomic Research of Tunisia)- for sharing their knowledge and expertise..
- All the TMT participants for their interest and valuable contribution provided during the TMT.

Supplementary File 1: Additional details for Fig. 5.

REFERENCES

Beavis, W., & Mahama, A. A. (2023). Multi Environment Trials: Linear Mixed Models. In Quantitative Genetics for Plant Breeding. Iowa State University Digital Press.

Breese, E.L. (1989). Regeneration and multiplication of germplasm resources in seed genebanks: the scientific background. IBPGR, Rome, Italy, pp. 69.

Brush, S.B. (2005). Genes in the field: on farm conservation of crop diversity. Lewis Publ., Boca Raton, USA.

Chung, Y. S., Choi, S. C., Jun, T. H., & Kim, C. (2017). Genotyping-by-sequencing: a promising tool for plant genetics research and breeding. Horticulture Environment and Biotechnology, 58(5), 425–431. https://doi.org/10.1007/S13580-017-0297-8/METRICS

Churchill, G. A., & Doerge, R. W. (1994). Empirical threshold values for quantitative trait mapping. Genetics, 138(3), 963–971. https://doi.org/10.1007/s11703-007-0022-y

COGENT (2018). A global strategy for the conservation and use of coconut genetic resources 2018-2028. (R. Bourdeix and A. Prades, compilers), Bioversity International, Montpellier, France.

Cronn, R., Knaus, B. J., Liston, A., Maughan, P. J., Parks, M., Syring, J. V., & Udall, J. (2012). Targeted enrichment strategies for next-generation plant biology. American Journal of Botany, 99(2), 291–311. https://doi.org/10.3732/ajb.1100356

Crop Genebank Knowledge Base. https://cropgenebank.sgrp.cgiar.org/index.php/home-mainmenu-71/about-this-site

Cross, R.J., & Wallace, A.R. (1994). Loss of genetic diversity from heterogeneous self-pollinating genebank accessions. Theor. Appl. Genet., 88, 885-890.

De Boef, W.S., Subedi, A., Peroni, N., Thijssen, M., &O'Keeffe, E. (2013). Community Biodiversity Management. Routledge Ltd., Oxon., UK.

Delahaye, C., & Nicolas, J. (2021). Sequencing DNA with nanopores: Troubles and biases. PLoS ONE, 16(10). https://doi.org/10.1371/JOURNAL.PONE.0257521

Elshire, R. J., Glaubitz, J. C., Sun, Q., Poland, J. A., Kawamoto, K., Buckler, E. S., & Mitchell, S. E. (2011). A Robust, Simple Genotyping-by-Sequencing (GBS) Approach for High Diversity Species. PLoS ONE, 6(5), e19379. https://doi.org/10.1371/journal.pone.0019379

FAO (2011). Introduction tothe International Treaty on Plant Genetic Resources for Food and Agriculture. http://www.fao.org/3/a-i2631e.pdf

Gale, J.S., & Lawrence, M.J. (1984). The decay of variability. In Crop Genetic Resources: Conservation and Evaluation (J.H.W. Holden & J.T. Williams, Eds.), George Allen & Unwin Ltd., Hert., UK, pp. 77-101.

Geibelid, J., Reimer, C., Weigendid, S., Weigend, A., Pookid, T., &Simianer, H. (2021). How array design creates SNP ascertainment bias. https://doi.org/10.1371/journal.pone.0245178

Giani, A. M., Gallo, G. R., Gianfranceschi, L., &Formenti, G. (2020). Long walk to genomics: History and current approaches to genome sequencing and assembly. Computational and Structural Biotechnology Journal, 18, 9–19. https://doi.org/10.1016/J.CSBJ.2019.11.002

Goodwin, S., McPherson, J. D., & Richard McCombie, W. (2016). Coming of age: ten years of next-generation sequencing technologies. https://doi.org/10.1038/nrg.2016.49

Heywood, V.H., & Dulloo, M.E. (2005). *In situ* conservation of wild plants: a critical review of best practisers. IPGRI Tech. Bull., 11, IPGRI, Rome, Italy.

Hu, T., Chitnis, N., Monos, D., & Dinh, A. (2021). Next-generation sequencing technologies: An overview. Human Immunology, 82(11), 801–811.

https://doi.org/10.1016/j.humimm.2021.02.012 Ikegawa, S. (2012). A Short History of the Genome-Wide Association Study: Where We Were and Where We Are Going. Genomics & Informatics, 10(4), 220. https://doi.org/10.5808/GI.2012.10.4.220

Kapli, P., Yang, Z., & Telford, M. J. (2020). Phylogenetic tree building in the genomic age. Nature Reviews Genetics, 21(7), 428–444. https://doi.org/10.1038/s41576-020-0233-0

Li, H., & Durbin, R. (2024). Genome assembly in the telomere-to-telomere era. Nature Reviews Genetics. https://doi.org/10.1038/s41576-024-00718-w

Miller, M. R., Dunham, J. P., Amores, A., Cresko, W. A., & Johnson, E. A. (2007). Rapid and cost-effective polymorphism identification and genotyping using restriction site associated DNA (RAD) markers. Genome Research, 17(2), 240–248. https://doi.org/10.1101/GR.5681207

PacBio. (2024). Long-read sequencing myths: debunked. Part 1- HiFi sequencing. https://www.pacb.com/blog/long-read-sequencing-myths-debunked-part-1-hifi-sequencing

Pritchard, J. K., Stephens, M., & Donnelly, P. (2000). Inference of population structure using multilocus genotype data. Genetics, 155(2), 945–959. https://doi.org/10.1111/j.1471-8286.2007.01758.x

Sims, D., Sudbery, I., Ilott, N. E., Heger, A., & Ponting, C. P. (2014). Sequencing depth and coverage: Key considerations in genomic analyses. Nature Reviews Genetics, 15(2), 121–132. https://doi.org/

Sohn, J., & Nam, J.-W. (2016). The present and future of de novo whole-genome assembly. Briefings in Bioinformatics, bbw096. https://doi.org/10.1093/bib/bbw096

Song, L., Huang, W., Kang, J., Huang, Y., Ren, H., & Ding, K. (2017). Comparison of error correction algorithms for Ion Torrent PGM data: application to hepatitis B virus. Scientific Reports 2017 7:1, 7(1), 1–11. https://doi.org/10.1038/s41598-017-08139-y

Thomas, D. C., Haile, R. W., & Duggan, D. (2005). Recent Developments in Genomewide Association Scans: A Workshop Summary and Review. The American Journal of Human Genetics, 77(3), 337–345. https://doi.org/10.1086/432962

Wilson, E.O. (1988). Biodiversity. National Acad. of Sc., Washington D.C. Natl. Acad. Press, USA. https://doi.org/10.17226.989.

Yu, J., Pressoir, G., Briggs, W. H., Vroh Bi, I., Yamasaki, M., Doebley, J. F., McMullen, M. D., Gaut, B. S., Nielsen, D. M., Holland, J. B., Kresovich, S., & Buckler, E. S. (2006). A unified mixed-model method for association mapping that accounts for multiple levels of relatedness. Nature Genetics, 38(2), 203–208. https://doi.org/10.1038/ng1702

Zhang, J., Liu, F., Reif, J. C., & Jiang, Y. (2021). On the use of GBLUP and its extension for GWAS with additive and epistatic effects. G3 Genes|Genomes|Genetics, 11(7). https://doi.org/10.1093/g3journal/jkab122