ECOBREED workshop

Soybean advanced genotyping

ecobreed
IMPROVING CROPS

Funded by European Union
Horizon 2020
Grant agreement No 771367

1. What is genotyping?

Genotyping is the process that detects genetic differences that can lead to major changes in phenotype. Genotyping examines DNA sequencesusing biological assays and compares this to reference sequences. It reveals the alleles that have been inherited from the parents. It has a vast range of uses across basic scientific research, medicine, and agriculture.

Traditionally genotyping is the use of DNA sequences to define biological populations by use of molecular tools. It does not usually involve defining the genes of an individual. Current methods of genotyping include the use of molecular markers using restriction enzyme, polymerase chain reaction (PCR), or DNA sequencing. Genotyping is important in research of genes and gene variants associated with traits.

Understanding genetic variation has a particular benefit in the agricultural world, where trait selection in plants and livestock has been used for centuries to increase performance, yield and quality.

While traditional selective breeding involved purely observational methods (selecting only plants or animals with superior phenotypic traits, such as size or strength, for breeding), modern selective breeding relies heavily on molecular techniques, including SNP genotyping.

Selective breeding pressures have generated animal breeds and plant varieties with more desirable phenotypes and changes to specific genomic regions associated with these phenotypes. Detecting these functionally relevant genetic changes helps us to understand which particular genes and sequences are associated with specific traits. This is useful for designing new and more intelligent breeding programs.

Online Tutorials

- Brief description of what is genotyping? What does genotyping mean? Genotyping meaning, definition and explanation; https://www.youtube.com/watch?v=NxqqRNt3ub0
- High-throughput genotyping solutions for challenges in commercial plant breeding presented by T. Osborn director of Molecular Breeding technology (LCG group); https://www.youtube.com/watch?v=zvwQjez6AIQ
- SNP Genotyping Technologies by CD Genomics; https://www.youtube.com/watch?v=plWYBLy9OaM

Further reading

- Gaj P, Maryan N, Hennig EE, Ledwon K, Paziewska A. 2012. Pooled sample-based GWAS: A cost-effective alternative for identifying colorectal and prostate cancer risk variants in the Polish population. PLOS One 7(4):e35307.
- Kwok P-Y, Chen X. 2003. Detection of single nucleotide polymorphisms. Curr. Issues. Mol. Biol 5:43-60.
- Rathnayake I, Hargreaves M, Huygens F. 2011. SNP diversity of Enterococcus faecalis and Enterococcus faecium in a Southeast Queensland waterway, Australia, and associated antibiotic resistance gene profiles, BMC Microbiol. 11(1):201.
- Reddy MPL, Wang H, Liu S, Bode B, Reed JC, Steed RD, She JX. 2011. Association between type 1 diabetes and GWAS SNPs in the southeast US Caucasian population. Genes. Immun. 12(3):208-212.
- Sengstake S, Bablishvili N, Schuitema A, Bzekalava N, Abadia E, de Beer J, Bergval I. 2014. Optimizing multiplex SNP-based data analysis for genotyping of Mycobacterium tuberculosis isolates, BMC Genomics, 15(1):572.
- Visscher PM, Wray NR, Zhang Q, Sklar P, McCarthy MI, Brown MA, Yang J. 2017. 10 Years of GWAS Discovery: Biology, Function, and Translation. Am. J. Hum. Genet. 101(1):5-22.

2. Overview of quantitative genetics

The characters studied by Mendel have only two distinct possibilities. The pea seeds were either round or wrinkled, with a color either yellow or green, etc. The so call qualitative traits have a variability which is not continuous (discrete variation).

Other traits do not fall into discrete classes. Rather, when a segregating population is analyzed, a continuous distribution of phenotypes is found. An example is seed size. When a large seed line is crossed with a small seed line, the seed size of theF1 is intermediate to the two parents. Furthermore, when the F1 plants are inter- mated, the distribution of seed size in the F2 ranges from the small to large size. The distribution resembles the bell-shaped curve for a normal distribution.

These types of traits are called continuous traits and cannot be analyzed in the same manner as discontinuous traits. Because continuous traits are often measured and given a quantitative value, they are often referred to as quantitative traits, and the area of genetics that studies their mode of inheritance is called quantitative genetics.

Many important agricultural traits such as crop yield, weight gain, plant height, color intensity, etc. are quantitative traits, and much of the pioneering research into the modes of inheritance of these traits was performed by agricultural geneticists.

Due to the continuous distribution of phenotypic values, quantitative genetics must employ many other statistical methods (such as the effect size, the mean and the variance) to link phenotypes (attributes) to genotypes. Some phenotypes may be analyzed either as discrete categories or as continuous phenotypes, depending on the definition of cut-off points, or on the metric used to quantify them. Mendel himself had to discuss this matter in his famous paper, (Mendel, 1866). “Versuche über Pflanzen Hybriden”. *Verhandlungen Naturforschender Verein in Brünn*] especially with respect to his peas attribute tall/dwarf, which actually was “length of stem”. Analysis of quantitative trait loci, or QTL, is a more recent addition to quantitative genetics, linking it more directly to molecular genetics.

The reasons for the normal distribution of quantitative traits have been justified by the Johannsen and Nilsson-Ehle experiments. The first demonstrated that variation in some traits is due to environmental variation which affects the phenotype expression while the second attributed variation to the effect of multigenes with addictive actions. These several pairs of genes are called **quantitative trait loci (QTL)** and this is called **polygenic inheritance** or the **multiple-factor hypothesis**.

2.1 Heritability

Heritability measures the fraction of phenotype that can be attributed to genetic variation.

The phenotype can be modelled as the sum of genetic and environmental effects:

Phenotype (P) = Genotype (G) + Environment (E).

Likewise, the phenotypic variance in the trait $\sigma^2(P)$ is:

$$\sigma^2(P) = \sigma^2(G) + \sigma^2(E) + 2\,\mathrm{Cov}(G,E).$$

In a planned experiment Cov(G,E) can be controlled and held at 0. In this case, heritability is:

$h_b^2 = \sigma^2(G)/ \sigma^2(P)$.

h_b^2 is the broad-sense heritability. This reflects all the genetic contributions to a population's phenotypic variance including additive, dominant, and epistatic (multi-gene interactions).

A particularly important component of genetic variance is the additive variance, $\sigma^2(A)$, which is the variance due to the average effects (additive effects) of thealleles. Since each parent passes a single allele per locus to each offspring, parent- offspring resemblance depends upon the average effect of single alleles. Additive variance represents, therefore, the genetic component of variance responsible for parent-offspring resemblance. The additive genetic portion of the phenotypic variance is known as Narrow-sense heritability and is defined as: $h^2 = \sigma^2(A)/ \sigma_n^2(P)$.

Further Reading

- Davies SW, Scarpino SV, Pongwarin T, Scott J, Matz MV. 2015. Estimating Trait Heritability in Highly Fecund Species G3: Genes, Genomes, Genetics 5: 2639-2645. https://doi.org/10.1534/g3.115.020701.
- Wray N, Visscher P. 2008. Estimating trait heritability. Nature education, 1(1), 29.

3. Genetic resources and their conservation

To run any breeding program and/or to undertake genotypic characterization, it is important to use a wide range of genetic variation. Hence, it is imperative to properly conserve germplasm and understand the taxonomic relationships and cross-ability potential in the selected plant genetic resources. The importance of conserving genetic diversity is to have sources of variability for using in breeding programs. This is ever more and more important due to the increase and range of available technologies as well as the urgency of increasing rates of genetic erosion and overall loss of biodiversity. When developing a comprehensive plant breeding program, this should address not only the selection of the species, but also the choice of the correct parental material and a program of crossing activity.

During the 1920s and 1930s, the Russian botanist, Nikolai Ivanovich Vavilov, undertook a series of excursions around the world. He noted that species diversity was not randomly distributed but was higher for each crop in a particular region of theworld. He identified eight centers of origin of cultivated species. These are still in use and are as follows: (1) Mexico-Guatemala, (2) Peru-Ecuador-Bolivia, (2A) Southern Chile, (2B) Southern Brazil, (3) Mediterranean, (4) Middle East, (5) Ethiopia, (6) Central Asia, (7) Indo-Burma, (7A) Siam-Malaya-Java, and (8) China.

Within each center of diversity, the variation found is higher for specific cultivated crops. Later studies distinguished between centers of origin i.e., where levels of diversity are high, with the presence of wild progenitors (from where the species probably originated) and centers of diversity, i.e., where the species variation is high, but without the presence of those crop wild relatives. For example, the Middle East is the center of origin for wheat while Ethiopia is a center of wheat diversity, with only tetraploid species present. Harlan (1971) then identified larger areas where domestication took place and, considering their magnitude, defined those areas as “not-centre”.

Plant breeding and genetic resources

Public awareness of genetic resources conservation has been steadily increasing, especially as a key component of sustainable agriculture as well as mitigating the impact of large-scale monocultures and the consequential effects produced by these human activities. Despite this recognition of the importance of crop genetic resources, which has been well known since plant breeders started their activities (XIX century), but the awareness of its importance and the danger of erosion and disappearance is limited almost exclusively to scientists directly involved. Publicconcerns are generally focused on the extinction of some endangered minor wild species, but nobody at that level worries about the dramatic impact that shrinking of genetic variation in crop plants and their wild relatives may have on future food production. Since a limited component of modern societies, except farmers, areconcerned with agriculture itself. Even if crop genetic resources represent the basisof agricultural development, they provide an enormous reservoir of useful genes and gene complexes that endow plants to cope with evolving resources and habitats. Without the availability of a reserve of variation and different alleles able to let the crops react differently to the different needs, which could be resistance to both biotic and abiotic factors, but also the needs of new products different in color, shape,test, etc., the breeder would not have the starting

material to do his work.

Genes for adaptability are the main resource for affording challenges of a changing environment. Unfortunately, the destruction of natural ecosystems has severely reduced the genetic variability in wild species, whereas the replacement of local varieties with improved ones has virtually eliminated landraces. Most of the abundant genetic resources available a few decades ago have been lost forever. The concern for genetic resources started during the 1950s when scientists started travelling world-wide and especially in developing countries when it was pointed out what was happening with the cultivation of newly acquired crop land and the spreading of modern uniform varieties. Problems shifted from technical, to financial and finally to political grounds. A wide movement took place; the onset of recombinant DNA technology, at the beginning of the 80s shifted the attention from Genetic Resources to advanced biotechnology as possible source of variation, and to some extent alleviated the concerns. The limits of manipulations based on molecular and tissue culture techniques have been recognized and it is becoming clear thatonly an integrated approach between traditional and advanced techniques will produce the best results. It is obvious, in fact, that useful genes have to be identified, their role in metabolic pathways clarified, and their membership to gene families and/or complexes ascertained, and this requires the availability of the widest genetic variation, for as many species as possible. For this reason, seed companies started to be interested both in genetic resources and biotechnology.

Further reading

- De Vicente (Ed). 2006. DNA banks – providing novel options for genebanks? Topical Views in Agricultural Biodiversity. Bioversity International, Rome, Italy.
- Dulloo ME, Labokas J, Iriondo JM, Maxted N, Lane A, Laguna E, Jarvis A, Kell SP. 2008. Genetic reserve location and design. In Conserving Plant Genetic Diversity in Protected Areas (Eds J.M. Iriondo, N. Maxted and M.E. Dulloo). Cabi International.
- Engels JMM, Maggioni L, Maxted N, Dulloo ME. 2008. Complementing in situ conservation with ex situ measures. In Conserving Plant Genetic Diversity in Protected Areas (Eds J.M. Iriondo, N. Maxted and M.E. Dulloo). Cabi International.
- Eyzaguirre PB, Linares OF. 2004. Home Gardens and Agrobiodiversity. Smithsonian Books, Washington, DC, USA.
- Fowler C, Hodgkin T. 2004. Plant genetic resources for food and agriculture: assessing global availability. Annu. Rev. Environ. Resour. 29, 143-179.
- Harlan JR, De Wet JMJ. 1971. Toward a rational classification of cultivated plants. Taxon, 20 (4), 509-517.
- Laliberté B. 1997. Botanic garden seed banks/genebanks worldwide, their facilities, collections and networks. Bot. Gardens Cons. News, 2(9), 18-23.
- Maxted N, Iriondo JM, Dulloo ME, Lane A. 2008. The integration of PGR conservation with protected area management. In Conserving Plant Genetic Diversity in Protected Areas (Eds Iriondo J.M., Maxted N., Dulloo M.E.). CAB International.
- Pence VC, Sandoval JA, Villalobos VM, Engelmann F. 2002. In vitro collecting techniques for germplasm conservation. IPGRI Technical Bulletin No.7. Bioversity International, Rome, Italy.
- Scarascia-Mugnozza GT, Perrino P. 2002. History of ex situ conservation and the use of PGR. In Managing Plant Genetic Diversity (J.M.M. Engels, V.R. Rao, A.H.D. Brown and M.T: Jackson Eds). IPGRI, Italy.

4. Trait selection in agriculture

The development of improved varieties has made a major contribution to the increased productivity and quality of plants used for their food, feed, fiber or esthetic value. Selection of an appropriate variety is one of the key decisions that a farmer has to make since the variety will define the limits of performance that can beachieved in any environment. Plant breeding has been a part of agriculture since humans first selected one type of plant or seed in preference to another instead of randomly taking what nature provided. This led to elimination of undesirable characters such as seed dormancy and shattering. Preferential selection to meet human needs resulted in a broad range of cultivated types within species. The overallobjective of plant breeding is to improve those characteristics of a species thatcontribute to its economic value. The part of plant having economic value may beleaf, stem, root, flower, seed or fruit. Selection can be made for direct improvement ofthe plant part or for the characters that are related to reliability of production, harvest- ability and marketability. There is a vast list of characteristics considered by plant breeders. Traits of primary importance for plant breeding common for many species are: yield – the amount of production per unit area; resistance to pests and diseases
– genetic resistance is the most effective mean of biological control; seed composition – the value of seed may be influenced by its chemical composition, content, quality and nutritive value; forage quality – animal productivity is related to the quality of the forage consumed; tolerance to mineral stress – where crops are grown on soils with undesirable characteristics; tolerance to environmental stress – temperature and water extremes can cause major reductions in crop productivity; adaptability to mechanization – modification of certain characteristics enables highly mechanized crop production and harvesting.

Variety development involves the application of knowledge provided by a number of scientific disciplines and their integration in an effective program. Central disciplines of plant breeding are considered agronomy, horticulture, genetics and in the last decade different 'omics' technologies (e.g. genomics, proteomics, metabolomics). The method by which a crop is produced and utilized determines the characters that are important for selection and the conditions under which the characters should be evaluated. Knowledge of the inheritance of a character is basic, which as well as qualitative and quantitative genetics contribute to the understanding of plant behavior in breeding process. Today's plant breeding utilizes as a foundation the genetic principles initiated by the classic investigations of Gregor Mendel rediscovered in the early 1900s. Mendel employed the sound scientific principle of reducing a complex question to its component parts for study and then bringing the parts together for the final conclusions. He was able to accurately describe inheritance mechanisms based on assumptions of paired units and random transmission of the units from parent to progeny. His laws of segregation and independent assortment are as valid today as at the time when they were discovered. Since then, a vast number of plant inheritance studies has occurred in some of which the trait of interest was simply inherited. For a trait to be defined as simply inherited a single gene or tightly grouped cluster of linked genes, inherited asa unit, must be responsible. For any given population structure simply inherited traits segregate among progeny at expected Mendelian ratios and include those traits in which a completely dominant phenotype can be qualitatively scored. Quantitative traits are determined by the action and interaction of two or more

genes or gene × environment interactions and can be defined as one whose genetic component does not follow strict Mendelian inheritance. Many traits of primary interest to breeders are genetically quite complex (e.g. yield, dormancy, nutritional traits) and provide significant analytical challenges requiring dense linkage maps and well replicatedsets of phenotypic data. The application of molecular genetics is an important contribution and addition to plant breeding programs. Because agronomic traits are quantitatively inherited, quantitative trait loci (QTL) discovery represents a valuable tool for enhancing yield and yield stability of crop production while maximizing its sustainability. Genomics approaches will allow more efficient discovery and manipulation of QTL and will become increasingly important for coping with the challenges faced by crop production. QTL studies allow us to investigate cause-effect relationships between traits. A better understanding of the QTL that underlines these traits would provide new momentum for more targeted selection programs based on marker assisted selection which is already an important component of different breeding programs, particularly in the private sector.

Further reading

- Ceoloni C, Forte P, Kuzmanovic L, Tundo S, Moscetti I, De Vita P,·Virili ME,·D'Ovidio R. 2017. Cytogenetic mapping of a major locus for resistance to Fusarium head blight and crown rot of wheat on *Thinopyrum elongatum* 7EL and its pyramiding with valuable genes from a *Th. ponticum* homoeologous arm onto bread wheat 7DL. Theor. Appl. Genet. 130: 2005-2024.
- Ceoloni C, Kuzmanovic L, Forte P, Virili ME, Bitti A. 2015. Wheat-perennial Triticeae introgressions: major achievements and prospects. In: Molnár-Láng M, Ceoloni C, Doležel J (Eds.), Alien Introgression in Wheat - Cytogenetics, Molecular Biology, and Genomics. Springer, pp. 273–313.
- Ceoloni C, Kuzmanović L, Gennaro A, Forte P, Giorgi D, Grossi MR, Bitti A. 2014.Genomes, chromosomes, and genes of perennial triticeae of the genus Thinopyrum: the value of their transfer into wheat for gains in cytogenomic knowledge and 'precision' breeding. In: Tuberosa R, Graner A, Frison E. (Eds.), Advances in Genomics of Plant Genetic Resources. Springer, Dordrecht, The Netherlands, pp. 333–358.

- Forte P, Virili ME, Kuzmanović L, Moscetti I, Gennaro A, D'Ovidio R, Ceoloni C. 2014. A novel assembly of *Thinopyrum ponticum* genes into the durum wheat genome: pyramiding Fusarium head blight resistance onto recombinant lines previously engineered for other beneficial traits from the same alien species. Mol. Breed 34: 1701–1716.
- Gennaro A, Forte P, Carozza R, Savo Sardaro ML, Ferri D, Bitti A, Borrelli GM, D'Egidio MG, Ceoloni C. 2007. Pyramiding different alien chromosome segments in durum wheat: feasibility and breeding potential. Isr. J. Plant Sci. 55: 267-276.
- Gennaro A, Forte P, Panichi D, Lafiandra D, Pagnotta MA, D'Egidio MG, Ceoloni C. 2012. Stacking small segments of the 1D chromosome of bread wheat containing major gluten quality genes into durum wheat: transfer strategy and breeding prospects. Mol. Bree 30: 149-167.
- Kuzmanović L, Gennaro A, Benedettelli S, Dodd IC, Quarrie SA, Ceoloni C. 2014. Structural-functional dissection and characterization of yield-contributing traits originating from a group 7 chromosome of the wheatgrass species *Thinopyrum ponticum* after transfer into durum wheat. J. Exp. Bot. 65, 509–525.
- Kuzmanović L, Ruggeri R, Able JA, Bassi FM, Maccaferri M, Tuberosa R, De Vita P, Rossini F, Ceoloni C. 2018. Yield performance of chromosomally engineered durum wheat-*Thinopyrum ponticum* recombinant lines in a range of contrasting rain-fed environments. Field Crops Res. 228: 147-157.
- Meglič V. 2013. Morpho-agronomic traits. Maloy S.R., Hughes K.T. (ed.) Brenner's encyclopedia of genetics. Second ed. vol. 4, 475-477.
- Moore G. 2009. Early stages of meiosis in wheat and the role of *Ph1*. In: Muehlbauer G, Feuillet C (eds) Genetics and Genomics of the Triticeae. Plant Genetics and Genomics: Crops and Models. Pp237-252.
- Singh RP, Huerta-Espino J, Rajaram S, Crossa J. 1998. Agronomic effects from chromosome translocations 7DL.7Ag and 1BL.1RS in spring wheat. Crop Sci. 38: 27–33.
- Vitellozzi F, Ciaffi M, Dominici L, Ceoloni C. 1997. Isolation of a chromosomally engineered durum wheat line carrying the common wheat *Glu-D1d* allele. Agronomie 17: 413-419.

Molecular procedures and tools

DNA extraction methods

As molecular marker technology is evolving into a more and more valuable tool for creating new plant cultivars (Kikuchi et al. 2017; Dayteg at al. 2017), it is important to provide good quality, high yield genetic material and a consistent method for its extraction. This can serve as a basis for further molecular genetic analysis (Abdel Latif and Osman 2017), for instance PCR and real time PCR analysis, Southern blotting, restriction enzyme digestion, NGS-based applications, etc. A number of different commercial kits for DNA extraction are available on the market nowadays, differing in isolation technology, sample type and amount; time needed per run, elution volume, DNA yield and potential downstream applications. Most commonly, these kits are based on solid-phase nucleic acid purification (Tan and Yiap 2009) andperformed by using a spin column, operated under centrifugal force (Gjerse et al. 2009). That results in fast and efficient DNA purification in comparison to conventional methods, such as cetyltrimethylammonium bromide (CTAB) or sodium dodecyl sulfate (SDS) methods (Tan and Yiap 2009). When preparing plant tissue forthe DNA extraction it is very important to consider:

1. the type of plant tissue;
2. that we are selecting the young and healthy parts of the plant tissue;
3. the amount of starting plant material (too much is not always better);
4. how is the starting material stored (fresh, -20^{o}C, -80^{o}C);
5. the appropriate homogenization method (time of grinding is also important);
6. the use of extra additions to the homogenization buffer if needed.

It is important to point out that optimized DNA extraction methodology is crucial step when obtaining suitable template for further genotyping applications. Moreover, required time and cost of a particular method should not be ignored, especially when dealing with a high number of samples.

Online Tutorials

- DNA extraction by Lab Center at DNALC; http://labcenter.dnalc.org/labs/dnaextraction/dnaextraction_d.html
- Plant genomic DNA extraction by Rahul Patharkar; https://www.youtube.com/watch?v=p2RARQj0X9Y
- Plant genomic DNA isolation by Disha Lifesciences Ltd; https://www.youtube.com/watch?v=9-olhcH4Nl4

Further reading

- Abdel Latif A, Osman G. 2017. Comparison of three genomic DNA extraction methods to obtain high DNA quality from maize. Plant Methods. 13(1).
- Azmat MA, Khan IA, Cheema HMN, Rajwana IA, Khan AS, Khan AA. 2012. Extraction of DNA suitable for PCR applications from mature leaves of *Mangifera indica* L. J Zhejiang Univ-Sci B (Biomed Biotechnol). 13(4):239–243.
- Dayteg C, Tuvesson S, Merker A, Jahoor A, Kolodinska-Brantestam A. 2017. Automation of DNA marker analysis for molecular breeding in crops: practical experience of a plant breeding company. Plant Breed. 126:410–415.
- Derlink M, Pipan B, Pavlovčič P, Jones, LE, Meglič V, Symondson WO, Virant-Doberlet M. 2014C. Characterization of eleven polymorphic microsatellite markers for leafhoppers of the genus Aphrodes (Hemiptera: Cicadellidae). Conserv Genet Resour. 6(4):933–935.
- Gjerse DT, Hoang L, Hornby D. 2009. RNA purification and analysis: sample preparation, extraction, chromatography. 1st ed. Weinheim, Germany: Wiley-VCH.
- Kikuchi T, Kasajima I, Morita M, Yoshikawa N. 2017. Practical DNA markers to estimate apple (*Malus X domestica* Borkh.) skin color, ethylene production and pathogen resistance. J Hortic. 4(4):211.
- Manoj K, Tushar B, Sushama C. 2007. Isolation and purification of genomic DNA from black plum (*Eugenia jambolana* Lam.) for analytical applications. Int. J. Biotechnol Biochem. 3(1):49–55.
- Maras M, Pipan B, Šuštar Vozlič J, Todorović V, Đurić G, Vasić M, Meglič V. 2015. Examination of genetic diversity of Common bean from the Western Balkans. J. Am. Soc. Hortic. Sci. 140(4):308–316.
- Mikulic Petkovsek M, Slatnar A, Stampar F, Veberic R. 2010. The influence of organic/integrated production on the content of phenolic compounds in apple leaves and fruits in four different varieties over a 2-year period. J Sci Food Agric. 90:2366–2378.
- Pipan B, Šuštar Vozlič J, Meglič V. 2013. Genetic differentiation among sexually compatible relatives of Brassica napus L. Genetika. 45(2):309–327.
- Pipan B., Zupančič M., Blatnik E., Dolničar P., Meglič V. 2018. Comparison of six genomic DNA extraction methods for molecular downstream applications of apple tree (*Malus X domestica*). Cogent food & agriculture, p, no. 4, p. 1-10.
- Rusjan D, Pelengić R, Pipan B, Or E, Javornik B, Štajner N. 2015. Israeli germplasm: phenotyping and genotyping of native grapevines (*Vitis vinifera* L.). Vitis. 54:87–89.
- Sahu SK, Thangaraj M, Kathiresan K. 2012. DNA extraction protocol for plants with high levels of secondary metabolites and polysaccharides without using liquid nitrogen and phenol. ISRN Mol. biol. [6 p.]. DOI: 10.5402/2012/205049.
- Sinkovič L, Pipan B, Meglič V, Kunstelj N, Nečemer M, Zlatić E, Žnidarčič D. 2017. Genetic differentiation of Slovenian sweet potato varieties (*Ipomoea batatas*) and effect of different growing media on their agronomic and nutritional traits. Ital. J. Agron. 12(4):350– 356.

Polymerase Chain Reaction (PCR) – main principles

PCR is a widely used method in molecular biology for making copies of a specific DNA segment. Most PCR methods amplify DNA fragment lengths of between 0.1and 10-kilo base pairs (kbp). By PCR, a single copy (or more) of a DNA sequence is exponentially amplified to generate thousands to millions copies of that particular DNA segment. The main steps of PCR are represented in Fig. 8.1. During thedenaturation step (94-96°C), the DNA double helix is separated by breaking the hydrogen bonds between the nucleotides to allow copying of the given fragment. At the annealing step, the temperature is lowered to 50-65°C (depending on the primers used) to allow binding of primers to regions flanking the DNA stretch to be amplified. Finally, at the extension step (68-78°C), DNA polymerase synthesizes the new DNA strand based on the DNA template.

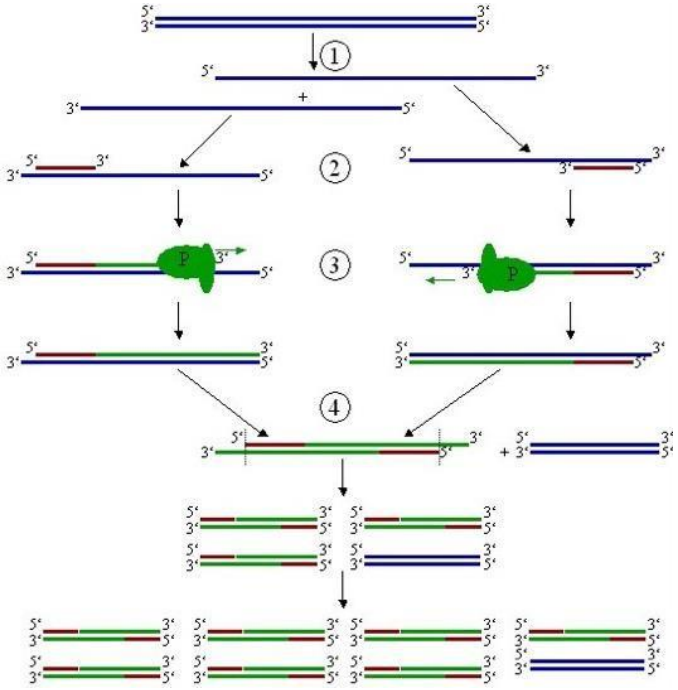

Fig. 8.1. Scheme of a polymerase chain reaction and main stages: initialisation and denaturation (1), annealing (2), extension (3) and final elongation (4) (in blue, the DNA to amplify; in red, primers; in green, the newly synthesised DNA strand; P, DNA polymerase).

4.1. PCR mixture components

4.1.1. Primers – characteristics and principles for design

At the beginning of a PCR, the fragment to be amplified must be flanked by short single-stranded DNA sequences named primers ("forward" - P_f and "reverse" - P_r) which serves as a probe for initiation of the reaction (Fig. 8.2a). Primers are indispensable for PCR because many DNA polymerases i.e. enzymes that catalyse DNA replication, cannot begin an *ex-novo* synthesis of a new strand. Each primer of a primer pair is oriented towards the other one with its 3'-OH terminals and must respect the principle of complementarity between the nucleotides in the site that flanks the stretch of DNA to be amplified (Fig. 8.2b).

Fig. 8.2. Orientation of primers with respect to the DNA sequence to amplify (a) and principle of complementarity between the primer and the DNA sequence (b).

The extension of the two primers during PCR always occurs in the 5' - 3' direction, as indeed happens during the replication *in-vivo*. The primers serve to form an initial duplex with the DNA template filament and to provide the free 3'-OH site that the DNA polymerase I recognizes as suitable for starting replication. Design of primer sequence is one of the most critical factors for PCR and in general, should produce primer pairs where each primer is 17-24 base pair long, has around 50% GC content,and the two sequences must not allow the formation of internal H bonds e.g. primer- dimers or hairpin loops (Fig. 8.3).

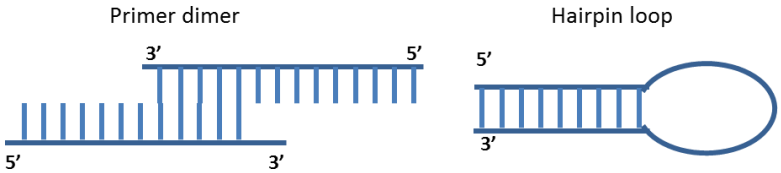

Fig. 8.3. Examples of secondary structure that can form within a primer pair due to the excess of complementarity between the sequences of the two member primers.

Annealing temperatures (T_a) of the two primers in a pair must be similar and lower than the "melting" temperature of the DNA template sequence. To determine the final T_a for a given primer pair, Wallace's rule should be initially followed: for each A/T, and each G/C base in the primer sequence, 2 and 4°C are added, respectively, and summed all together [T_a = 2 * (A + T) + 4 * (G + C)]. This rule serves to provide an indication of the pairing temperature of a certain primer. The higher the T_a, the more specific is the pairing. The lower the T_a, the more possibility there is for a non-specific pairing, which would lead to amplification of products other than the one of interest.

Different types of molecular markers depend on the specific primer sequences (RAPD, AFLP, SSR, EST, etc., see chapter below).

4.1.2. DNA polymerase I

The enzyme *Taq* polymerase, normally used in PCRs, was identified and isolated from the DNA of the bacterium *Thermus aquaticus*) and is stable at high temperatures necessary for the denaturation of double-stranded DNA. Its optimal functioning is at 68-72°C, depending on the type of *Taq* and of the producer. The action of this enzyme is based on formation of the phosphodiester bond between a 3'-OH terminal of a primer nucleotide and the alpha phosphate on the free 5'- triphosphate (dATP, dGTP, dCTP, dTTP), with pyrophosphate release (PPi). The enzyme has a 5'- 3' exonuclease activity, with no 3'-5' exonuclease activity.

4.1.3. DNA template

The amount of DNA template often determines the specificity of the reaction, the more DNA the more specific is the reaction. Moreover, the quality of the DNA should be sufficient for normal DNA polymerase activity, hence protocols for DNA extraction should eliminate all excessive amounts of proteins, sugars and other compounds that could interfere with the DNA polymerase.

4.1.4. $MgCl_2$

Optimal concentration of divalent cations (Mg^{++}) is necessary and critical for the *Taq* action. It is important to pay attention that the reaction solution does not contain an excess of chelating agents, such as EDTA, which could capture magnesium, and make it unavailable for the DNA polymerase.

It is essential to apply good laboratory practice to avoid contamination of the PCR mixture by unwanted exogenous DNA, which could represent the biggest problem a successful outcome of the experiment. Therefore, the area where a PCR mixture is prepared must be different from that where the DNA samples are managed (ideally under a hood used exclusively for PCR preparation). Pipettes to be used for PCR mixture preparation should not be used for any other solvent or DNA/RNA, and reagents divided in small aliquots. Frequent change of gloves, thorough cleaning of work surfaces and instrumentation, immediate closure of the tubes immediately after use should also make part of the routine practice.

4.2. Main PCR types

Standard PCR - use of a common DNA polymerase and conditions of amplification.

Long-range PCR - refers to the amplification of DNA lengths up to 40kbp by using specific methods and reagents. Usually, DNA polymerases with proofreading activity are used and longer extension times. It is used for the analysis/cloning of long DNA fragments, amplification of particularly long gene sequences, or the analysis of chromatin rearrangements.

Asymmetric PCR – refers to preferential amplification of one of the two DNA strands of the template. An excess of one of the two primers is used, leading to the exponential amplification of the targeted strand. This type of PCR is used in sequencing and hybridisation.

High fidelity PCR – uses *Taq* polymerases with high ability of accurate replication of the desired template i.e., with combined low non-incorporation rates and proofreading activity. Its main applications are sequencing of *in-vitro* amplified material, cloning, protein expression or gene studies, SNP analysis, or RNA analysis by RT-PCR (see ahead).

Hot-start PCR - use of DNA polymerase type which can be gradually activated only after exposure to 95°C for 2-15 minutes (depending on the manufacturer). This PCR reduces non-specific amplification during the initial set up stages of the PCR and allows for PCR mixture preparation at room temperature.

Touch down PCR – programmed to perform cycles in which the annealing temperature is progressively lowered during the various phases of the PCR, from an initial higher value with respect to the expected T_m, to a value lower than the expected T_m. It aims to reduce non-specific DNA amplification.

Multiplex PCR – use of more than 1 primer pair in a single PCR mixture to produce amplicons of various sizes. In this way, multiple polymorphic sequences or genes can be targeted in a single reaction and reduce the time of the overall analyses. Annealing temperatures for each of the primer sets must be optimised to work correctly within a single reaction. In addition, the amplicon profiles i.e., band sizes must be different enough to enable easy visualisation by gel electrophoresis.

Nested PCR – PCR that involves two pairs of primers used in two successive PCR reactions, where the second pair of primers amplifies the PCR product of the first one. It increases the specificity of the reaction and reduces the non-specific DNA amplification.

Real Time PCR - also known as quantitative PCR (qPCR), represents a method of simultaneous amplification and quantification of DNA. It is commonly used to determine with high precision the number of target DNA copies in the sample. The qPCR uses fluorescent dyes, such as Sybr Green to measure the amount of amplified product in real time.

Reverse transcriptase Polymerase chain reaction (RT-PCR) for RNA – technique that combines the reverse transcription of RNA into DNA (i.e. cDNA) and amplification of specific DNA targets using standard PCR. It is primarily used to measure the amount of a specific RNA. RT-PCR can be used without qPCR, to enable molecular cloning, sequencing or simple detection of RNA.

Online tutorials

- Basis of a PCR by Applied Biological Materials – abm; https://www.youtube.com/watch?v=matsiHSuoOw
- Primer3 web site is a widely used program for designing PCR primers, but also for designing hybridization probes and sequencing primers; http://primer3.ut.ee/

Further reading

- BatchPrimer3 v1.0, A high throughput web application for PCR and sequencing primer. https://probes.pw.usda.gov/batchprimer3/index.html
- GrainGenes online tool for designing genome-specific primers in polyploidy species. https://wheat.pw.usda.gov/GG3/node/248

5. Markers and Molecular Tools

The assessment of genetic diversity within and between populations is evaluated using morphological, biochemical, and molecular characterization and evaluation:

1. *Morphological* characterization does not require expensive technology, but large amounts of land are often required for these experiments, making it possibly more expensive than molecular assessment. These traits are often susceptible to phenotypic plasticity; conversely, this allows assessment of diversity in the presence of environmental variation.
2. *Biochemical* analysis is based on the separation of proteins into specific banding patterns. It is a fast method which requires only small amounts of biological material. However, only a limited number of enzymes are available and thus, the resolution of diversity is limited.
3. *Molecular* analyses comprise a large variety of DNA molecular markers, which can be employed for analysis of variation. Different markers have differentgenetic qualities (they can be dominant or co-dominant, can amplify anonymous or characterized loci, can contain expressed or non-expressed sequences, etc.).

The concept of genetic markers is not a new one; Gregor Mendel employed phenotype-based genetic markers in his experiments. Later, phenotype-based genetic markers for *Drosophila melanogaster* led to the founding of the theory of genetic linkage, occurring when particular genetic loci or alleles for genes are inherited jointly. The limitations of phenotype-based genetic markers led to the development of DNA-based markers, i.e. molecular markers. A molecular marker can be defined as ‘’*a genomic locus, detected through probe or specific starters (primer) which, by virtue of its presence, distinguishes unequivocally the chromosomic trait which it represents as well as the flanking regions at the 3’ and 5’ extremities’’*.

Molecular markers may or may not correlate with phenotypic expression of a genomic trait. They offer numerous advantages over conventional, phenotype-based alternatives as they are stable and detectable in all tissues regardless of growth, differentiation, development, or defense status of the cell. Additionally, they are not confounded by environmental, pleiotropic and epistatic effects.

An ideal molecular marker should possess the following features: (i) be polymorphic and evenly distributed throughout the genome; (ii) provide adequate resolution of genetic differences; (iii) generate multiple, independent and reliable markers; (iv) be simple, quick and inexpensive; (v) need small amounts of tissue and DNA samples; (vi) link to distinct phenotypes; and (vii) require no prior information about the genome of an organism. Nevertheless, no molecular marker presents all the listed advantages.

The different methods of molecular assessment differ from each other with respect to important features such as genomic abundance, level of polymorphism detected, locus specificity, reproducibility, technical requirements and cost. Depending on the need, modifications in the techniques have been made, leading to a second generation of advanced molecular markers.

Genetic or DNA based marker techniques such as Restriction Fragment Length Polymorphism (RFLP), Random Amplified Polymorphic DNA (RAPD), Simple Sequence Repeats (SSR) and Amplified Fragment Length Polymorphism (AFLP) are now in common use for ecological, evolutionary, taxonomical, phylogenic and genetic studies of plant sciences. These techniques are well established and their advantages and limitations have been well documented (Ayad et al. 1995; Agarwal etal. 2008). The recent development of high-throughput sequencing technology provides the possibility of analyzing high numbers of samples over smaller periods of time.

Molecular Markers

Molecular markers work by highlighting differences (polymorphisms) within a nucleic sequence between different individuals. These differences include insertions, deletions, translocations, duplications and point mutations. They do not, however, encompass the activity of specific genes.

In addition to being relatively impervious to environmental factors, molecular markers have the advantage of: (i) being applicable to any part of the genome (introns, exons, and regulation regions); (ii) not possessing pleiotropic or epistatic effects; (iii) being able to distinguish polymorphisms which do not produce phenotypicvariation and finally, (iv) being some of them co-dominant. The different techniques employed are based either on restriction-hybridisation of nucleic acids or techniques based on Polymerase Chain Reaction (PCR), or both. In addition, the different techniques can assess either multi-locus or single-locus markers. Multi-locus markers allow simultaneous analyses of several genomic loci, which are based on the amplification of casual chromosomic traits through oligonucleic primers with arbitrary sequences. These types of markers are also defined as *dominant* since it is possible to observe the presence or the absence of a band for any locus, but it is not possible to distinguish between heterozygote (a/-) conditions and homozygote for the same allele (a/a). By contrast, single-locus markers employ probes or primers specific to genomic loci and can hybridize or amplify chromosome traits with well- known sequences. They are defined as co-dominant since they allow discrimination between homozygote and heterozygote loci.

5.1. Non-PCR-Based Techniques

5.1.1. Restriction-Hybridization Techniques

Molecular markers based on restriction-hybridization techniques were employed relatively early in the field of plant studies and combined the use of *restriction endonucleases* and the hybridization method (Southern 1975). Restriction endonucleases are bacterial enzymes able to cut DNA, identifying specific palindrome sequences and producing polynucleotidic fragments with variable dimensions. Any changes within sequences (i.e. point mutations), mutations between two sites (i.e. deletions and translocations), or mutations within the enzyme site, can generate variations in the length of restriction fragment obtained after enzymatic digestion.

RFLP and Variable Numbers of Tandem Repeats (VNTRs) markers are examples of molecular markers based on restriction-hybridisation techniques. In RFLP, DNA polymorphism is detected by hybridising a chemically-labelled DNA probe to a

Southern blot of DNA digested by restriction endonucleases, resulting in a differential DNA fragment profile. The RFLP markers are relatively highly polymorphic, co-dominantly inherited, highly replicable, and allow the simultaneous screening of numerous samples. DNA blots can be analysed repeatedly by stripping and re-probing (usually eight to ten times) with different RFLP probes. Nevertheless, this technique is not very widely used as it is time-consuming, involves expensive and radioactive/toxic reagents and requires large quantities of good quality genomic DNA.Moreover, the pre-requisite of prior sequence information for probe construction contributes to the complexity of the methodology. These limitations led to the development of a new set of less technically complex methods known as PCR-based techniques.

5.2. Markers Based on Amplification Techniques (PCR-Derived)

The use of this kind of marker has been exponential, following the development by Mullis et al. (1986) of the Polymerase Chain Reaction (PCR). This technique consists of the amplification of several discrete DNA products, derived from regions of DNA which are flanked by regions of high homology with the primers. These regions must be close enough to one another to permit the elongation phase.

The use of random primers overcame the limitation of prior sequence knowledge for PCR analysis and being applicable to all organisms, facilitated the development of genetic markers for a variety of purposes. PCR-based techniques can further be subdivided into two subcategories: (1) arbitrarily primed PCR-based techniques or sequence non-specific techniques; and (2) sequence targeted PCR-based techniques. Based on this, two different types of molecular markers have been developed: RAPD and AFLP.

5.2.1. Random Amplified Polymorphic DNA (RAPD)

RAPDs were the first PCR-based molecular markers to be employed in genetic variation analyses. RAPD markers are generated through the random amplification of genomic DNA using short primers (decamers), separation of the obtained fragments on agarose gel in the presence of ethidium bromide and finally, visualisation under ultraviolet light. The use of short primers is necessary to increase the probability that, although the sequences are random, they are able to find homologous sequences suitable for annealing. DNA polymorphisms are then produced by “rearrangements or deletions at or between oligonucleotide primer binding sites in the genome” (Williams et al. 1991). As this approach requires no prior knowledge of the genome analysed, it can be employed across species using universal primers. The major drawback of this method is that the profiling is dependent on reaction conditions which can vary between laboratories; even a difference of a degree in temperature is sufficient to produce different patterns. Additionally, as several discrete loci are amplified by each primer, profiles are not able to distinguish heterozygous from homozygous individuals (Bardakci 2001). Arbitrarily Primed Polymerase Chain Reaction (AP-PCR) and DNA Amplification Fingerprinting (DAF) are independently developed methodologies, which are variants of RAPD. For AP-PCR, a single primer, 10–15 nucleotides long, is used and involves amplification for initially two PCR cycles at low stringency. Thereafter, the remaining cycles are carried out at higher stringency by increasingthe annealing temperatures.

5.2.2. Amplified Fragment Length Polymorphism (AFLP)

To overcome the limitation of reproducibility associated with RAPD, AFLP technology was developed by the Dutch company, Keygene (Vos et al. 1995). This

method is based on the combination of the main analysis techniques: digestion of DNA through restriction endonuclease enzymes and PCR technology. It can beconsidered an intermediate between RFLPs and RAPDs methodologies as it combines the power of RFLP with the flexibility of PCR-based technology.

The primer pairs used for AFLP usually produce 50–100 bands per assay. The number of amplicons per AFLP assay is a function of the number selective nucleotides in the AFLP primer combination, the selective nucleotide motif, GC content, and physical genome size and complexity. AFLP generates fingerprints of any DNA regardless of its source, and without any prior knowledge of the DNA sequence. Most AFLP fragments correspond to unique positions on the genome and hence can be exploited as landmarks in genetic and physical mapping. The technique can be used to distinguish closely related individuals at the sub-species level and can also map genes.

The origins of AFLP polymorphisms are multiple and can be due to: (i) mutations of the restriction site which create or delete a restriction site; (ii) mutations of sequences flanking the restriction site, and complementary to the extension of the selective primers, enabling possible primer annealing; (iii) insertions, duplications or deletions inside amplification fragments. These mutations can cause the appearance/disappearance of a fragment or the modification (increase or decrease) of an amplified-restricted fragment.

5.2.3. Sequence Specific PCR Based Markers

A different approach to arbitrary PCR amplification consists of the amplification of target regions of a genome through specific primers. With the advent of high-throughput sequencing technology, abundant information on DNA sequences for the genomes of many plant species has been generated. Expressed Sequence Tags (EST) of many crop species have been generated and thousands of sequences have been annotated as putative functional genes using powerful bioinformatics tools. ESTs are single-read sequences produced from partial sequencing of a bulk mRNA pool that has been reverse transcribed into cDNA. EST libraries provide a snapshot of the genes expressed in the tissue at the time of, and under the conditions in which,they were sampled (Bouck and Vision 2007). Despite these advantages, however, EST-SSRs are not without their drawbacks. One of the concerns with SSRs ingeneral is the possibility of null alleles, which fail to amplify due to primer sitevariation, do not produce a visible amplicon. Because the cDNA from which ESTs arederived lack introns, another concern is that unrecognised intron splice sites could disrupt priming sites, resulting in failed amplification. Lastly, as EST-SSRs are located within genes, thus more conserved across species, they may be less polymorphic than anonymous SSRs. Although the use of EST possesses these limitations, several features of EST sequence libraries make them a valuable resource for conservation and evolutionary genetics. ESTs are an inexpensive source for identifying gene-linked markers with higher levels of polymorphism, which can also be applied to closely related species. EST libraries are also a good starting point for developing tools to study gene expression such as microarrays or quantitative PCR assays.

5.2.4. Microsatellite-Based Marker Technique

Microsatellites or Simple Sequence Repeats (SSR) are sets of repeatedsequences found within eukaryotic genomes (Morgante and Olivieri 1993). These consist of sequences of repetitions, comprising basic short motifs generally between2 and 6

base-pairs long. Polymorphisms associated with a specific locus are due to the variation in length of the microsatellite, which in turn depends on the number of repetitions of the basic motif. Variations in the number of tandemly repeated units are mainly due to strand slippage during DNA replication where the repeats allow matching via excision or addition of repeats (Schlotterer and Tautz 1992). As slippage in replication is more likely than point mutations, microsatellite loci tend to be hypervariable. Microsatellite assays show extensive inter-individual length polymorphisms during PCR analysis of unique loci using discriminatory primers sets.

Microsatellites are highly popular genetic markers as they possess: co-dominant inheritance, high abundance, enormous extent of allelic diversity, ease of assessing SSR size variation through PCR with pairs of flanking primers and high reproducibility. However, the development of microsatellites requires extensive knowledge of DNA sequences, and sometimes they underestimate genetic structure measurements, hence they have been developed primarily for agricultural species, rather than wild species. Initial approaches were principally based on hybridisation techniques, whilst more recent techniques are based on PCR (Gupta and Varshney 2000). Major molecular markers based on assessment of variability generated by microsatellites sequences are: STMSs (*Sequence Tagged Microsatellite Site*), SSLPs (*Simple Sequence Length Polymorphism*), SNPs (*Single Nucleotide Polymorphisms*), SCARs (*Sequence Characterised Amplified Region*) and CAPS (*Cleaved Amplified Polymorphic Sequences*).

5.3. Single Nucleotide Polymorphisms (SNPs)

Single nucleotide variations in genome sequence of individuals of a population are known as SNPs. SNPs are the most abundant molecular markers in the genome. They are widely dispersed throughout genomes with a variable distribution among species. The SNPs are usually more prevalent in the non-coding regions of the genome. Within the coding regions, when an SNP is present, it can generate either non-synonymous mutations that result in an amino acid sequence change (Sunyaev et al. 1999), or synonymous mutations that do not alter the amino acid sequence. Synonymous changes can, however, modify mRNA splicing, resulting in phenotypic differences. Improvements in sequencing technology and an increase in the availability of the increasing number of EST sequences have made analysis of genetic variation possible directly at the DNA level.

Many SNP genotyping analyses are based on allele-specific hybridisation, oligonucleotide ligation, primer extension or invasive cleavage (Sobrino et al. 2005). Genotyping methods, including DNA chips, allele-specific PCR and primer extension approaches based on SNPs, are particularly attractive for their high data throughput and for their suitability for automation. They are used for a wide range of purposes, including rapid identification of crop cultivars and construction of ultra-high-density genetic maps.

5.4. Markers Based on Other DNA than Genomic DNA

There are also other highly informative approaches used to study genetic variation based on organelle microsatellite sequence detection; in fact, due to their uniparental mode of transmission, chloroplast (cpDNA) and mitochondrial genomes (mtDNA) exhibit different patterns of genetic differentiation compared to nuclear alleles (Provan et al. 1999; Breidenbach et al. 2019). Consequently, in addition to nuclear microsatellites, marker techniques based on chloroplast and mitochondrial

microsatellites have also been developed. The cpDNA, maternally inherited in most plants, has proved to be a powerful tool for phylogenetic studies. Due to increasing numbers of recent examples of intra-specific variation observed in cpDNA, there is additional potential for within-species genetic variation analysis. CpDNA has been preserved well within the genome, and consequently has been employed widely for studying plant populations through the use of PCR-RFLP and PCR sequencing approaches. They are also employed in the detection of hybridisation/introgression (Bucci et al. 1998), in the analysis of genetic diversity and in obtaining the phylogeography of plant populations.

Mitochondrial DNA in plants, in contrast, has been demonstrated to be an unsuitable tool for studying phylogenesis and genetic diversity, being quantitatively scarce. At the nuclear level, another type of sequence employed largely for studying genetic diversity is ribosomal RNA (rRNA). Ribosomal RNA genes are placed on the specific chromosomal loci *Nor* and organised in tandem repeats which can be repeated up to thousands of times. Since some regions of rRNA are well preservedin eukaryotes, it represents a very useful phylogenetic tool. Conversely, other regionssuch as the "Internal Transcriber Spacers" (ITS) are so variable that they can beused to analyse polymorphism at the intra-specific level.

5.5. Transposable Elements-Based Molecular Markers

Although transposon insertions can have deleterious effects on host genomes, transposons are considered important for adaptative evolution, and can beinstrumental in acquiring novel traits (Miller et al. 1997; May and Dellaporta 1998; Girard and Freeling 1999). Retrotransposons have so far received little attention in the assessment of genetic diversity, despite their contribution to genome structure, size, and variation. Additionally, their dispersion, ubiquity and prevalence in plant genomes provide an excellent basis for the development of a set of marker systems, to be used alone or in combination with other markers, such as AFLPs and SSRs. Retrotransposon-based molecular analysis relies on amplification using a primer corresponding to the retrotransposon and a primer matching a section of the neighbouring genome. To this type of class of molecular markers belong: *Sequence-Specific Amplified Polymorphism* (S-SAP), *Inter-Retrotransposon Amplified Polymorphism* (IRAP), *Retrotransposon-Microsatellite Amplified Polymorphism* (REMAP), *Retrotransposon-Based Amplified Polymorphism* (RBIP) and finally, *Transposable Display* (TD).

5.6. RNA-Based Molecular Markers

Studies of mechanisms which control genetic expression are essential to better understand biological responses and developmental programming in organisms. PCR-based marker techniques such as cDNA-SSCP, cDNA-AFLP and RAP-PCR are used for differential RNA studies, using selective amplification of cDNA.

5.7. Real-Time PCR

Real-time polymerase chain reaction is a laboratory technique based on the polymerase chain reaction, amplifying and simultaneously quantifying a targetedDNA molecule (Heid et al. 1996). It enables both detection and quantification (as absolute number of copies or relative amount when normalised to DNA input or additional normalising genes) of a specific sequence in a DNA sample. The procedure follows the general principle of polymerase chain reaction; its key featureis that the amplified DNA is quantified as it accumulates in the reaction in *real time* after each amplification

cycle. Two common methods of quantification are: (i) the use of fluorescent dyes that intercalate with double-stranded DNA and (ii) modified DNA oligonucleotide probes that fluoresce when hybridised with a complementary DNA. The major advantage of this technique consists in its sensitivity and speed due to the system of detection (spectrophotometric respect to ethidium bromide) and the quick changes of temperature. Real-time PCR is, therefore, particularly suitable for molecular markers based on PCR amplifications. In fact, the number of conservation and phylogenetic studies are now increasingly using real-time PCR for assessmentof genetic variation (Pagnotta et al. 2009).

5.8. Diversity Arrays Technology (DArT)

DArT is a generic and cost-effective genotyping technology. It was developed to overcome some of the limitations of other molecular marker technologies such as RFLP, AFLP and SSR (Akbari et al. 2006). DArT is an alternative method to time-consuming hybridisation-based techniques, typing simultaneously several thousand loci in a single assay. DArT is particularly suitable for genotyping polyploid species with large genomes, such as wheat. This technology generates whole-genome fingerprints by scoring the presence/absence of DNA fragments in genomic representations generated from samples of genomic DNA. DArT technology consists of several steps: (i) complexity reduction of DNA; (ii) library creation; (iii) the microarray of libraries onto glass slides; (iv) hybridisation of fluoro-labelled DNA onto slides; (v) scanning of slides for hybridisation signal and (vi) data extraction and analysis. DArT acts by reducing the complexity of a DNA sample to obtain a “representation” of that sample. The main method of complexity reduction used relies on a combination of restriction enzyme digestion and adapter ligation, followed by amplification. However, an infinite range of alternative methods can be used toprepare genomic representations for DArT analysis. DArT markers for a new species are discovered by screening a library of several thousand fragments from a genomic representation prepared from a pool of DNA samples that encompass the diversity of the species. The microarray platform makes the discovery process efficient because all markers on a particular DArT array are scored simultaneously. For each complexity reduction method, an independent collection of DArT markers can be assembled on a separate DArT array. The number of markers for a given species, therefore, is only dependent on: (i) the level of genetic variation within the species (orgene pool); and (ii) the number of complexity reduction methods screened.

5.9. New Generation of Sequencing Technology

The recent development of “high throughput sequencing” technologies makesDNA sequencing particularly important for conservation biology. These technologies have the potential to remove one of the major impediments to implementing genomic approaches in non-model organisms, including many of conservation relevance, i.e., the lack of extensive genomic sequence information. These technologies, in fact avoid the expense, complication, and biases associated with traditional clone-based sequencing by using direct amplification of DNA templates (Bentley 2006; Mardis 2008). The three pre-eminent technologies to be commercialized are 454 (Roche), Solexa (Illumina), and SOLiD (Applied Biosystems). The 454 sequencing is a pyrosequencing-based method that utilises emulsion PCR to achieve high throughput, parallel sequencing. Solexa’s sequencing-by-synthesis (SBS) approachis based on a simplified library construction method and reversible fluorescence termination

chemistry in the sequencing reaction, which produces 35-bp reads. Supported oligonucleotide ligation and detection (SOLiD) sequencing has some features in common with the other two technologies but, unlike the other two technologies, uses ligation-based sequencing technology. These new approaches to DNA sequencing enable the generation of 0.1–4 gigabases of DNA sequence in oneto seven days with reagent costs being between US$ 3,400 and 8,500. Due to the differences in fragment read lengths of sequencing, the target of each of these technologies is different: the shorter length and lower price per base of Solexa and SOLiD. This makes these approaches well suited to whole genome resequencing, where a novel genome sequence can be assembled and then compared to a reference sequence, that is, when the genome sequence of the species already exists. The 454 sequencing, on the other hand, with longer read lengths (soon to be upward of 400 bp per sequence) can also be used for obtaining the first glimpse of a species' genome or transcriptome.

5.10. Genotyping by sequencing

With the reduction of sequence cost and the speed up of the procedures next generation sequencing (NGS) technologies allow large-scale genome wide variation in populations to be obtained. Hence, genotyping by sequencing (GBS) has become popular to identify large scale variation in species both with and without a reference genome. Using GBS it is possible to identify thousands of single nucleotide polymorphism (SNP) markers, which can be used to analyse genetic variation within and between populations, and facilitate the analysis and dissection of complex traits, especially those involved in adaptive selection (Elshire 2011). GBS has two advantages (a) lower cost compared to the other techniques to identify SNPs in different species and crops and (b) it provides satisfactory results in the characterisation of germplasm, population studies and breeding.

Online Tutorials

- Markers Molecular/Genetic/DNA, Biochemical and Phenotypic by XploreBio; https://www.youtube.com/watch?v=Quk-Dh65iHY
- SNP (single nucleotide polymorphism) by XploreBio; https://www.youtube.com/watch?v=aaJkGFrWzFQ
- Introduction to Sequencing by Synthesis. Illumina Sequencing by Synthesis. By Illumina; https://www.youtube.com/watch?v=fCd6B5HRaZ8
- SSR Marker? Causes of SSR variation? Advantages, how to design SSR marker? by XploreBio; https://www.youtube.com/watch?v=iGN2tFCLPZ0
- genotyping mapping population using SSR marker by Genomics Lab; https://www.youtube.com/watch?v=dl0lTCBxNgE
- Genetic Markers | RAPD, RFLP, AFLP by Shomu's Biology; https://www.youtube.com/watch?v=JVM4LpCuT7g

Further reading

- Agarwal M, Shrivastava N, Padh H. 2008. Advances in molecular marker techniques and their application in plant sciences. Plant Cell Rep. 27: 617-631.
- Akbari M, Wenzl P, Caig V, Carling J, Xia L, Yang S, Uszinski G, Mohler V, Lehmensiek A, Kuchel H, Hayden M.J, Howes N, Sharp P, Vaughan P, Rathmell B, Huttner E, Kilian A. 2006. Diversity arrays technology (DArT) for high-throughput profiling of the hexaploid wheat genome. Theor. Appl. Genet. 113: 1409-1420.
- Ayad WG, Hodking A, Jaradat A, Rao VR. 1995. Molecular genetic techniques for plant genetic resources. IPGRI Workshop: Rome, Italy.
- Bardakci F. 2001. Random amplified polymorphic DNA (RAPD) markers. Turk. J. Biol. 25: 185-196.
- Bentley DR. 2006. Whole-genome re-sequencing. Curr. Opin. Genet. Dev. 16: 545-552.
- Bouck A, Vision T. 2007. The molecular ecologist's guide to expressed sequence tags. Mol. Ecol. 16: 907-924.
- Breidenbach N, Gailing O, Krutovsky KV. 2019. Development of novel polymorphic nuclear and chloroplast microsatellite markers in coast redwood (Sequoia sempervirens). Plant Genetic Res. 17(3), 293-297.
- Bucci G, Anzidei M, Madaghiele A, Vendramin GG. 1998. Detection of haplotypic variation and natural hybridization in halepensiscomplex pine species using chloroplast simple sequence repeat (SSR) markers. Mol. Ecol. 7: 1633-1643.
- Elshire RJ. 2011. A Robust, Si 571 mple Genotyping-by-Sequencing (GBS) Approach for High Diversity Species. Plos One 6, doi:10.1371/journal.pone.0019379.
- Girard L, Freeling M. 1999. Regulatory changes as a consequence of transposon insertion. Dev. Genet. 25: 291-296.
- Gupta PK, Varshney RK. 2000. The development and use of microsatellite markers for genetic analysis and plant breeding with emphasis on bread wheat. Euphytica 113: 163-185.
- Heid CA, Stevens J, Livak KJ, Williams PM. 1996. Real time quantitative PCR. Genome Res. 6: 986-994.
- Karp A, Seberg O, Buiatt, M. 1996. Molecular techniques in the assessment of botanical diversity. Ann. Bot., 78: 143-149.
- Mardis ER. 2008. The impact of next-generation sequencing technology on genetics. Trends Genet. 24: 133-141.
- May BP, Dellaporta SL. 1998. Transposon sequences drive tissue-specific expression of the maize regulatory gene R-s. Plant J. 13: 241-248.

- Miller W, McDonald J, Pinsker W. 1997. Molecular domestication of mobile elements. Genetica 100: 261-270.
- Morgante M, Olivieri AM. 1993. PCR-amplified microsatelites as markers in plant genetics. Plant J. 3: 175-182.
- Mullis KB, Faloona FA, Scharf SJ, Saiki SK, Horn GT, Erlich HA. 1986. Specific enzymatic amplification of DNA in vitro: the polymerase chain reaction. Cold Spring Harbor. Symp. Quant. Biol., 51: 263-273.
- Nadeem MA., Nawaz MA., Shahid M.Q., Doğan Y., Comertpay G, Yıldız M, Baloch FS. 2018. DNA molecular markers in plant breeding: current status and recent advancements in genomic selection and genome editing, Biotechnology & Biotechnological Equipment, 32:2, 261-285, DOI: 10.1080/13102818.2017.1400401.
- Pagnotta MA, Mondini L, Porceddu E. 2009. Quantification and organization of WIS2-1A and BARE-1 retrotransposons in different genomes of *Triticum* and *Aegilops* species. Molecular Genetics and Genomics. 282: 245-255.
- Provan J, Russell JR, Booth A, Powell W. 1999. Polymorphic chloroplast simple-sequence repeat primers for systematic and population studies in the genus *Hordeum*. Mol. Ecol. 8: 505-511.
- Reed D.H, Frankham R. 2003. Correlation between fitness and genetic diversity. Cons. Biol., 17: 230-237.
- Schlotterer C, Tautz D. 1992. Slippage synthesis of simple sequence DNA. Nucleic Acids Res. 20: 2211-2215.
- Southern E. 1975. Detection of specific sequences among DNA fragments separated by gel-electrophoresis. J. Mol. Biol. 98: 503.
- Spooner D, van Treuren R, de Vicente MC. 2005. Molecular Markers for Genebank Management. Bioversity International: Rome, Italy.
- Sunyaev S, Hanke J, Aydin A, Wirkner U, Zastrow I, Reich J, Bork P. 1999. Prediction of nonsynonymous single nucleotide polymorphisms in human disease-associated genes. J. Mol. Med. 77: 754-760.
- Vos P, Hogers R, Bleeker M, Reijans M, van de Lee T, Hornes M, Frijters A, Pot J, Peleman J, Kuiper M, Zabeau M. 1995. AFLP: a new technique for DNA fingerprinting. Nucleic Acids Res. 23: 4407-4414.
- Williams JGK, Kubelik AR, Livak K.J, Rafalski JA, Tingey SV. 1991. DNA polymorphisms amplified by arbitrary primers are useful as genetic markers. Nucleic Acids Res. 18: 6531-6535.

11. Methods/strategies for QTL identification

Quantitative Trait Locus (QTL) is a locus (DNA region) containing genetic factor(s) with additive effects on a specific polygenic quantitative trait i.e. a trait controlled by multiple genes and their interaction with the environment. Unlike monogenic traits, polygenic traits are not inherited according to classical Mendelian rules (discrete values), but their phenotypes vary along a continuous gradient of a bell curve. In crop plants, typical quantitative traits are yield, quality, flowering time or tolerance toabiotic stresses. By QTLs affecting only a portion of the variability of a given trait, the breeding process can be rather complicated.

Therefore, accurate identification of QTLs for agronomically useful traits is of paramount importance. To achieve progress in crop improvement for polygenic traits, mapping QTLs in the genome of crop species using molecular markers is indispensable. QTL mapping simply refers to finding an association between a genetic marker and a measurable phenotype. There are two main approaches to map QTLs:

1. **Linkage mapping method** - conventional mapping method, depends upon genetic recombination during the construction of mapping populations and has relatively low mapping resolution. The main steps include 1) creation of suitable mapping population (F_2, DH, BC, RIL, NIL), 2) selection of molecular markers (e.g., SSR, SNP) and construction of a linkage map, 3) genotyping of the mapping population and 4) linkage analysis with appropriate software package.
2. **Association or linkage disequilibrium (LD) mapping** - complementary to linkage mapping and takes advantage of historic recombination events accumulated over hundreds of generations, thus providing higher resolution and greater allele numbers. Owing to the dramatic reduction in costs of sequence technologies, association mapping has been conducted in many plants, including the major crops such as wheat, soybean, potato, maize, rice, sorghum, tomato, etc.

Steps in association mapping are 1) selection of association mapping panels from natural populations or germplasm collections with wide genetic diversity; 2) genotyping of the mapping population; 3) LD quantification by using molecular marker data and assessment of the population structure; 4) phenotyping of the panel and 5) correlation of phenotypic and genotypic data with an appropriate statistical approach to identify 'marker tags' positioned closely to the targeted trait.

There are two main types of association mapping: 1) candidate gene-based association mapping, which analyses polymorphisms of selected candidate genes and 2) genome-wide association mapping, which surveys genetic variation in the whole genome to associate allelic variation across the genome with various complex traits. For more details see Chapters on association mapping.

The key distinction between association and linkage mapping lies in whether recombination events occur in populations or families. Both of these methods share a consistent strategy for identifying molecular markers that are linked to QTL. It is also important to make a distinction between the terms linkage and LD, which are often confused. Linkage refers to the correlated inheritance of loci located on the same chromosome, whereas LD refers to the correlation between alleles in a population, but not necessarily on the same chromosome.

13.1 Types of mapping populations

Bi-parental – progeny obtained from the cross between two selected parents that have contrasting phenotypes for the trait of interest (F_2, BC, DH, RIL, NIL). They are typically used for classical linkage mapping, are easy to obtain, yet by combining the genomes of only two parents, a relatively narrow genetic base (low mapping resolution and large genetic intervals) is included that cannot adequately represent wider allelic diversity.

Multi-parent – emerged as next-generation mapping resources and combine diverse genetic founder contributions with high levels of recombination obtained. These populations derive from structured inter-mating between more than two well-characterised parents and maximize allelic diversity. In this way, the flaws of the bi-parental populations can be overcome by allowing the derivation of individuals which feature diverse levels and patterns of recombination and new genotype and haplotype combinations. The two most used multi-parent populations are the: (i) Nested Association Mapping (NAM) population - derived by crossing a single inbred parent to a successive collection of diverse inbred lines; and the (ii) Multi-parent Advanced Generation InterCrosses (MAGIC) population - developed by inter- crossing of multiple (typically four, eight or sixteen) parental lines in a balanced funnel crossing scheme.

Natural populations - collection of a sample population including elite cultivars, landraces, wild relatives, and exotic accessions are typically used for association mapping. Analysis of such populations include phenotyping and estimating broad-sense heritability of traits of interest, determining the genotypes of the population entries, quantification of the LD extent of the selected population and testing the associations between genotypes and phenotypes using appropriate statistical approaches.

13.2. Statistical analysis in genetic mapping and QTL detection

Linkage mapping analysis comprehend single-marker analysis (SIM), simple interval mapping (1 QTL at the time), composite interval mapping (CIM; identifies more QTL at the time, it is more precise).

Online Tutorials

- QTL Mapping Part 1 by Kristin Bishop-von Wettberg; https://www.youtube.com/watch?v=1JSw1gl3-RI
- QTL Mapping Part 2 by Kristin Bishop-von Wettberg; https://www.youtube.com/watch?v=Iu0SjECydK8
- Genome sequence and QTL identification for major agronomic traits of mung bean (Vigna radiata) by Suk-Ha Lee Seoul; https://www.youtube.com/watch?v=d17D5V0tqMo

Further reading

- Bernardo R. 2008. Molecular markers and selection for complex traits in plants: learning from the last 20 years. Crop Sci. 48: 1649-1664.
- Jansen RC, Tesson BM, Fu J, Yang Y, McIntyre LM. 2009. Defining gene and QTL networks. Curr. Opin. Plant Biol. 12: 241-246.

- Lincoln SE, Daly MJ, Lander ES. 1993. Mapping genes controlling quantitative traits using MAPMAKER/QTL version 1.1: a tutorial and reference manual. Whitehead Institute for Biometrical Research, Cambridge, Mass.
- McIntyre CL, Mathews KL, Rattey A, Chapman SC, Drenth J, Ghaderi M, Reynolds M, Shorter R. 2010. Molecular detection of genomic regions associated with grain yield and yield-related components in an elite bread wheat cross evaluated under irrigated and rainfed conditions. Theor. Appl. Genet. 120: 527-541.
- Yang J, Hu C, Hu H, Yu R, Xia Z, Ye X, Zhu J. 2008. QTLNetwork: mapping and visualizing genetic architecture of complex traits in experimental populations. Bioinformatics 24: 721-723.
- Parent B, Shahinnia F, Maphosa L, Berger B, Rabie H, Chalmers K, Kovalchuk A, Langridge P, Fleury, D. 2015. Combining field performance with controlled environment plant imaging to identify the genetic control of growth and transpiration underlying yield response to water-deficit stress in wheat. J. Exp. Bot. 66: 5481-5492.
- Quraishi UM, Pont C, Ain Q, Flores R, Burlot L, Alaux M, Quesneville H, Salse J. 2017. Combined genomic and genetic data integration of major agronomical traits in bread wheat (*Triticum aestivum* L.). Front. Plant Sci. 8:1843.
- Salvi S, Tuberosa R. 2015. The crop QTLome comes of age. Current Opinion Biotech. 32: 179-185.
- Tinker NA, Mather DE. 1995. MQTL: software for simplified composite interval mapping of QTL in multiple environments. J. Agric. Genomics 1: 1-5.
- Tondelli A, Francia E, Visioni A, Comadran J, Mastrangelo AM, Akar T, Al- Yassin A, Ceccarelli S, Grando S, Benbelkacem A, van Eeuwijk FA, Thomas WTB, Stanca AM, Romagosa I, Pecchioni N. 2014. QTLs for barley yield adaptation to Mediterranean environments in the 'Nure'×'Tremois' biparental population. Euphytica 197: 73-86.
- Utz HF, Melchinger AE. 1996. PLABQTL: a program for composite interval mapping of QTL. J. Quant. Trait Loci 2: 1-5.
- Utz HF. 2000. Introduction to PLABQTL. Institute of Plant Breeding, Seed Science and Population Genetics. University of Hohenheim, Germany. URL https://www.uni-hohenheim.de/fileadmin/einrichtungen/plant-breeding/software/pqintro.pdf.
- WGIN project website, provide genetic and molecular resources for wheat genetic stocks, mapping populations, molecular markers and marker technologies, trait identification and evaluation. http://www.wgin.org.uk/about.php.

14. Molecular-assisted selection - MAS

Principal goals of global plant breeding have typically aimed at improved yields, nutritional qualities, and other traits of commercial value. During the last thirty years, many studies have led to a rapid increase in knowledge of plant genome sequences and the physiological and molecular role of various genes, which have revolutionised molecular genetics and its own efficiency in genetic improvement programs (Nadeem et al. 2018).

Genetic mapping of major genes and quantitative traits loci (QTLs) for many important agricultural traits has increased the integration of biotechnology with the conventional breeding process. Therefore, DNA marker technology (see specific chapter 11) derived from research in molecular genetics and genomics, offers promising prospective for plant breeding. Owing to genetic linkage, molecular markers can be used to detect the presence of allelic variation in genes underling these traits. The use of molecular markers in plant breeding is called marker-assistedselection (MAS) and is a component of the new discipline of 'molecular breeding' (Collard and Mackill 2008). The volume of publications on the development and to a lesser extent application of markers for assisting plant breeding has increased dramatically during recent decades. The annual number of articles containing the term "marker assisted selection" surpassed 1000 in 2003 (Fig. 14.1.) (Xu and Crouch 2008) and probably an updated estimate would show further growth.

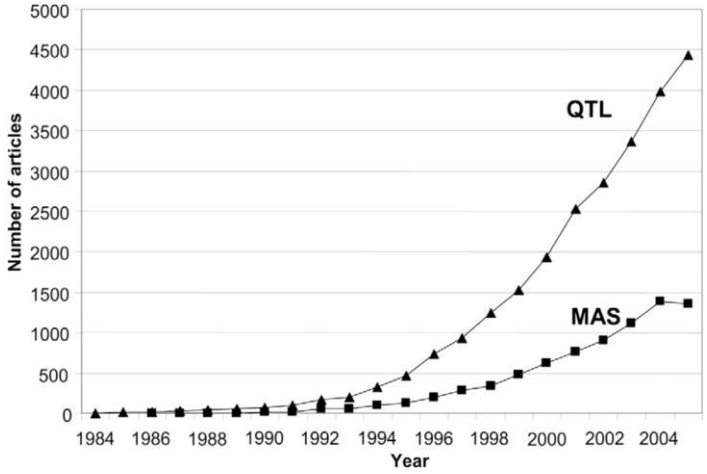

Fig. 14.1. The numbers of articles with the terms quantitative trait locus or quantitative trait loci (QTL) and marker-assisted selection (MAS) by years (1984-2005) from Google Scholar (4 Aug. 2007) (from Xu and Crouch, 2008).

Compared to classical breeding, MAS offers advantages in shortening the times of selection to obtain the desired phenotype based on the genotype identified with the markers (Collard et al. 2005), whose molecular profile is not influenced by environmental factors, but needs more specific and complex equipment and facilities. The first requirements for marker-assisted breeding (MAS) in plants should have: a) an appropriate marker system and reliable markers; b) quick DNA extraction and high throughput marker detection systems; c) knowledge of genetic linkage map and marker-trait association; d) quick and efficient data processing and management (Jiang 2013).

Plant breeders mostly use MAS for the identification of suitable dominant or recessive alleles across a generation and for identification of the most favourable individuals across the segregating progeny (Francia et al. 2005) (Fig. 14.2.).

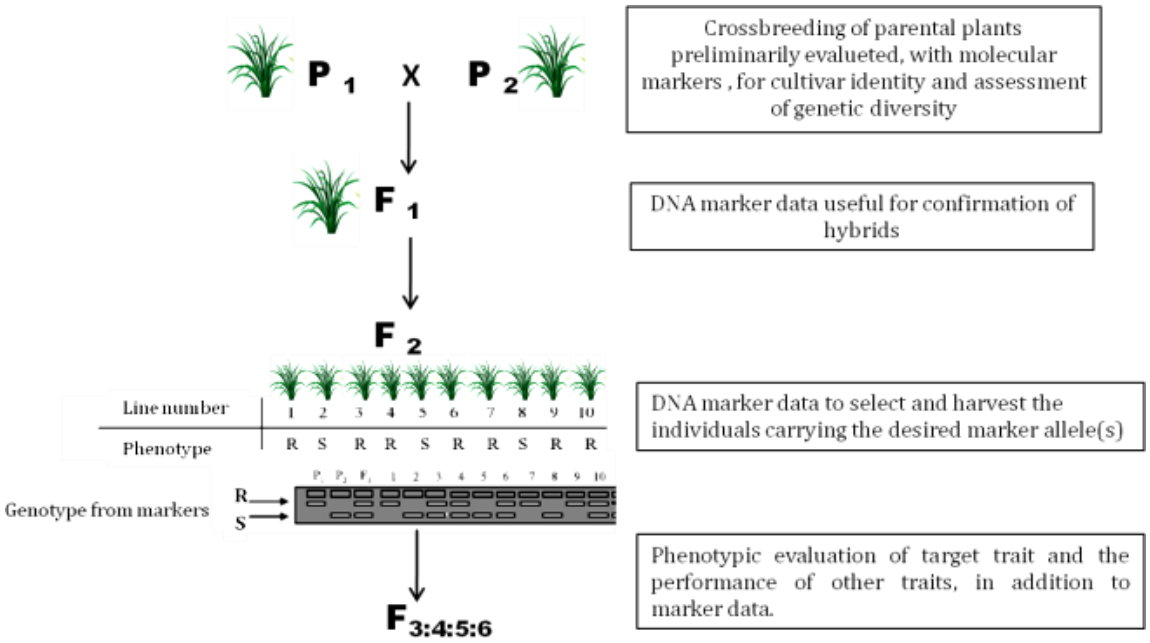

Fig. 14.2. An example of MAS approach. R= resistant, S= susceptible genotype.

An ideal DNA marker for MAS should be co-dominant (Fig. 14.3), evenly distributed throughout the genome, highly reproducible, at low cost, and having abilityto detect higher level of polymorphism (Nadeem et al. 2018).

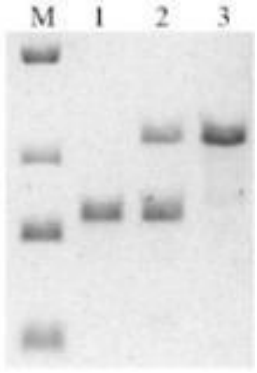

Fig. 14.3. Example of co-dominant marker. Lanes: M) DNA ladder; 1) dominant homozygote genotype; 2), heterozygote genotype; 3) recessive homozygote genotype.

For MAS it was suggested to limit the number of genes undergoing selection to three to four, if there are QTLs selected based on linked markers, and to five to six if there are known loci selected directly (Hospital 2003). The number of individuals in the screened population increases exponentially with the increase of target loci involved. The markers should be in the region of gene sequences or be close enoughto the gene/QTL of interest (<5cM) to ensure that only a minor proportion of the selected individuals will be recombinants. From the point of both effectiveness and efficiency, for a single QTL it is usually suggested to use two markers (i.e. flanking

markers) that are tightly linked to the gene/QTL of interest (Jiang 2013). Main schemes used for MAS are:

1. Selection based on mapped loci (QTL / genes): Marker-assisted backcrossing (MABC); Marker-assisted gene pyramiding (MAGP); Marker-assisted recurrent selection (MARS)
2. Selection with markers without map information of QTL / genes: Genomic selection (GS).

14.1. Marker-assisted backcrossing (MABC)

Special case of MAS in which breeding favourable alleles, to one or more loci, are transferred from a donor parent to an elite line through various cycles of backcross assisted selection with markers. Three general levels of marker-assisted backcrossing (MAB) have been described (Holland 2004) as:

1. 'Foreground selection': Selection for the allele most associated with the target gene (allele) provided by the parent donor (Fig. 14.4a)
2. 'Recombinant Selection': selection for the recurrent parent's alleles to the markers flanking the target gene, for reduced 'linkage drag' alongside the target gene (Fig. 15.4b)
3. 'Background selection': selection for alleles of the recurrent parent in the rest of the genome (optional) (Fig. 14.4c).

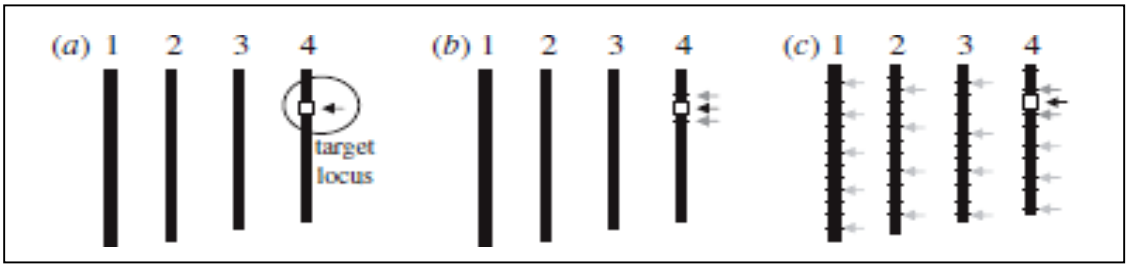

Fig. 14.4. Levels of selection during marker-assisted backcrossing. A hypothetical target locus is indicated on chromosome 4. (a) Foreground selection, (b) recombinant selection and (c) background selection (Collard and Mackill 2008).

Advantages of backcross assisted versus conventional MAS are: a) faster recovery of the recurrent parent's genome (often elite cultivar) and reduction of the "linkage drag" problem; b) like any type of MAS, it is not influenced by environmental factor; c) efficient selection of recessive alleles and individuals with event recombination near the target gene; d) best use of breeding program resources as the number of lines to keep per backcross cycle and the number of cycles to be performed are lower than in the traditional program.

14.2. Marker-assisted gene pyramiding

Marker-assisted gene pyramiding (MAGP) is one of the most important applications of DNA markers to plant breeding. This is a technique to enable transfer into a cultivar of QTLs/genes for single or multiple traits. This technique is mainly applied to increase the level of resistance to particular diseases and insects through the selection of two or more genes simultaneously (Nadeem et al. 2018).

MAS has been successfully applied to pyramid many desired genes in various crops (Ye et al. 2008; Gupta et al. 2010; Li et al. 2010; Wang et al. 2012).

14.3. Genomic selection (GS)

Genomic selection is based on markers without significant testing and without identifying *a priori* a subset of markers associated with the trait (Bernardo and Yu 2007). GS is a form of MAS, where marker effects across the entire genome (explaining entire phenotypic variation) are simultaneously estimated and used to calculate genomic estimated breeding values (GEBV) (Meuwissen et al. 2001; Heffner et al. 2009; Nakaya and Isobe 2012). Selection is then based on this breeding value rather than on a subset of significant markers, that are generally used in MAS (Gupta et al. 2010). While the MAS is commonly used for only major QTL/genes, so that breeding benefits are limited by the proportion of thegenotypic/phenotypic variance explained by markers associated with major QTLs, genomic selection (GS) could and should identify the whole of quantitative traits, that generally are controlled by a few major genes and many minor QTL/genes (Gupta et al. 2010).

Online Tutorials

- Description of Marker Assistant Selection by Dan Quiin for Shomu's Biology; https://www.youtube.com/watch?v=_OfDyT8E8dI
- Genomic Selection by Mark Sorrells Cornell University for Borlaug Global Rust Initiative; https://www.youtube.com/watch?v=s_FD7o5svZE
- Genomic Selection - Theory and Tools by Aaron Lorenz University of Nebraska -Lincoln for iPlant Collaborative; https://www.youtube.com/watch?v=PKc-IWKBD0c
- Eurofin web pages with links to Molecular Breeding and Genomic Technology video; https://www.eurofinsus.com/biodiagnostics/our-services/molecular-breeding/
- DNA Extraction and Marker Assisted Selection by TomatoLab; https://www.youtube.com/watch?v=yI8M9z4N4Y8

Further reading

- Bernardo R, Yu J. 2007. Prospects for genome wide selection for quantitative traits in maize. Crop Sci. 47:1082-1090.

- Collard BC, Jahufer MZ, Brouwer JB, Pang ECK. 2005. An introduction to markers, quantitative trait loci (QTL) mapping and marker-assisted selection for crop improvement: the basic concepts. Euphytica 142(1-2):169-196.
- Collard BCY, Mackill DJ. 2008. Marker-assisted selection: an approach for precision plant breeding in the twenty-first century. Phil. Trans. R. Soc. B 363, 557-572. doi:10.1098/rstb.2007.2170.
- Eathington SR, Crosbie TM, Edwards MD, Reiter RS, Bull JK. 2007. Molecular markers in a commercial breeding program. Crop Sci. 47:S154-S163.
- Francia E, Tacconi G, Crosatti C, Barabaschi D, Bulgarelli D, Dall'Aglio E, Valè G. 2005. Marker assisted selection in crop plants. Plant Cell Tiss Org. 82 (3):317-342.
- Gupta PK, Langridge P, Mir RR. 2010. Marker-assisted wheat breeding: present status and future possibilities. Mol Breed. 26(2):145-161.
- Heffner EL, Sorrells ME, Jannink JL. 2009. Genomic selection for crop improvement. Crop Sci 49:1-12.
- Holland JB. 2004. Implementation of molecular markers for quantitative traits in breeding programs-challenges and opportunities. In Proc. 4th Int. Crop Sci. Congress., Brisbane, Australia, 26 September-1 October.
- Hospital F. 2003. Marker-assisted breeding. In: H.J. Newbury (ed.), Plant molecular breeding. Blackwell Publishing and CRC Press, Oxford and Boca Raton, pp. 30-59.
- Jiang GL. 2013. Molecular Markers and Marker-Assisted Breeding in Plants. Plant Breeding from Laboratories to Fields, Chapter 3. http://dx.doi.org/10.5772/5258.
- Li X, Han Y, Teng W, Zhang S, Yu K, Poysa V, Anderson T, Ding J, Li W. 2010. Pyramided QTL underlying tolerance to Phytophthora root rot in mega-environment from soybean cultivar 'Conrad' and 'Hefeng 25'. Theor. Appl. Genet. 121, 651-658.
- Meuwissen T, Hayes BJ, Goddard ME. 2001. Prediction of total genetic value using genome-wide dense marker maps. Genetics 157, 1819-1829.
- Nadeem MA, Nawaz MA, Shahid MQ, Dogan Y, Comertpay G, Yıldız M, Hatipoglu R, Ahmad F, Alsaleh A, Labhane N, Ozkan H, Chung G, Baloch FS. 2018. DNA molecular markers in plant breeding: current status and recent advancements in genomic selection and genome editing. Biotechnology & biotechnological equipment, vol. 32, no. 2, 261- 285. https://doi.org/10.1080/13102818.2017.1400401.
- Nakaya A, Isobe SN. 2012. Will genomic selection be a practical method for plant breeding? Annals of Botany, DOI:10.1093/aob/mcs109.
- Ribaut JM, de Vicente MC, Delannay X. 2010. Molecular breeding in developing countries: challenges and perspectives. Current Opinion in Plant Biology, 13, 213-218. DOI 10.1016/j.pbi.2009.12.011.
- Wang X, Jiang G-L, Green M, Scott RA, Hyten DL. Cregan PB. 2012. Quantitative trait locus analysis of saturated fatty acids in a population of recombinant inbred lines of soybean. Mol. Breeding 30(2), 1163-1179. DOI 10.1007/s11032-012-9704-0.
- Xu Y, Crouch JH. 2008. Marker-Assisted Selection in Plant Breeding: From Publications to Practice. Crop Sci. 48, 391-407. doi: 10.2135/cropsci2007.04.0191.
- Ye G, Smith KF. 2008. Marker-assisted gene pyramiding for inbred line development: basic principles and practical guidelines. Int. J. Plant Breed. 2(1), 1-1