International Journal of Progressive Sciences and Technologies (IJPSAT)
ISSN: 2509-0119.

http://ijpsat.ijsht-journals.org
Vol. 26 No. 2 May 2021, pp.257-271

Effects Of Molecular Markers Consequences On Genotyping Errors

Ashraf M. Ward*[1], Miloud E Sweesi [2], Lutfi.L. Al-Mesilaty [3], Ali A. Ahmed[4],
Abdelatef A. Aswehli [1]. Abdulraouf M. Alkurdi[1], Giuma A. Elhafi [5]
Ismail M. Hdud[6] and Mohamed A. Benothman[7]

[1.]Department of Preventive Medicine, Faculty of Veterinary Medicine, University of Tripoli, Libya
[2.]The Faculty of Biotechnology, The University of Aljfara, Libya
[3.]Department of Microbiology, Faculty of Veterinary Medicine, Alzzytona University, Tarhona- Libya
[4.]Department of Pharmacology, Faculty of Veterinary Medicine, Alzzytona University, Tarhona- Libya
[5.]Department of Poultry and fish, Faculty of Veterinary Medicine, Alzzytona University, Tarhona- Libya
[6]Department of Pathology and Clinical Pathology, Faculty of Veterinary Medicine, University of Tripoli, Tripoli, Libya
[7.]Department of Animal Medicine, Faculty of Veterinary Medicine, University of Tripoli, Libya

Abstract – Genotype error can greatly reduce the power of a genetic study. For family data, genotype error can be assessed by examining marker data for non-Mendelian inconsistencies, closely linked markers for double recombination events, and consistency of duplicate genotypes. Genetic markers are widely used to determine the parentage of individuals in studies of mating systems, reproductive success, dispersals, quantitative genetic parameters and in the management of conservation populations. These markers are, however, imperfect for parentage analyses because of the presence of genotyping errors and undetectable alleles, which may cause incompatible genotypes (mismatches) between parents and offspring and thus result in false exclusions of true parentage. Highly polymorphic markers widely used in parentage analyses, such as microsatellites, are especially prone to genotyping errors. Molecular markers based on a hybridization reaction or PCR stage that detect DNA polymorphism are currently used in various fields of biology, including the study and preservation of genetic diversity, identification of individuals, phylogenetic, mapping of useful traits of quality and resistance to stress factors. , in the breeding process, biotechnology, etc. Before starting the experiment, researchers must determine what type of markers to use based on the following criteria: the variability and number of required markers, the need for their codominance, the corresponding requirements for the extracted DNA; practical - efficiency, reproducibility of analysis, the necessary appropriate technical support and cost. For a correct interpretation of the results of genotyping, one should take into account the fact that the use of any type of molecular markers is associated with a number of genotyping errors, the main ones being the loss of larger alleles, "zero" alleles, "stutter" alleles due to the peculiarities of Tag polymerase, and non-homology of amplified sequences of the same size (homoplasia). Researchers formulate defining conditions to reduce genotyping errors and reduce their impact on the final analysis. These include the quality and quantity of analyzed DNA, the level of technical capabilities and professionalism of the staff, since the human factor is defined as one of the main reasons for incorrect results; conducting pilot experiments for a comparative assessment of the theoretical and real error rate. The minimization of errors is achieved by assessing the capabilities of a particular type and screening markers; optimization of experimental methods, proper use of controls, replicates, and development of statistical approaches to identify errors. The trade-off between culling error-generating loci and increasing the potential of the retained loci to amplify the genetic signal can be different in different studies, but the main thing is that this signal is not lost for the sake of an "acceptable" level of error and the researchers develop empirical approaches to achieve the desired compromise.so that this signal is not lost for the sake of an "acceptable" level of error and researchers develop empirical approaches to achieve the desired compromise.so that this signal is not lost for the sake of an "acceptable" level of error and researchers develop empirical approaches to achieve the desired compromise.

Keywords – Molecular, Markers, Genotyping, Errors.

VARIETIES AND FEATURES OF MOLECULAR MARKERS

A simulation study was performed to investigate the effects of missing values, typing errors and distorted segregation ratios in molecular marker data on the construction of genetic linkage maps, and to compare the performance of three locus-ordering criteria (weighted least squares, maximum likelihood and minimum sum of adjacent recombination fractions criteria) in the presence of such effects. Molecular markers are nucleotide sequences and can be investigated through the polymorphism present between the nucleotide sequences of different individuals. Insertion, deletion, point mutations duplication and translocation are basis of these polymorphisms; however, they do not necessarily affect the activity of genes. An ideal DNA marker should be co-dominant, evenly distributed throughout genome, highly reproducible and having ability to detect higher level of polymorphism [10].

Progress in modern biology is largely based on the development and use of molecular genetic approaches, which are based on the analysis of DNA polymorphism, detected using various types of molecular markers. Molecular markers are currently used for genotyping when assessing genetic relationship / diversity between individuals, varieties, for creating genetic maps, mapping genes of interest, in the selection process (MAS, marker assisted selection), population genetics, phylogenetic studies, biotechnology, etc. [1-5].

The evolution of molecular markers is moving towards higher resolution, speed, simplicity and, if possible, lower analysis costs. Researchers [6] formulated requirements for DNA markers, according to which an ideal molecular marker must meet a set of characteristics: be highly polymorphic to reveal genetic diversity; have codominant inheritance, allowing you to determine the homozygous and heterozygous states of diploid organisms; the marker must be randomly and often distributed throughout the genome; its manifestation should be neutral (since the DNA sequences of any organism are neutral to external conditions and the procedures performed); the method should be simple and cheap to use; highly reproducible; should allow the exchange of results between laboratories. The presence of several types of markers is explained by the fact that none of them with a high degree meets the entire sum of the listed criteria. At the same time, it is obvious that the features of a particular type of molecular markers determine their more successful use for solving certain problems. Before starting the experiment, researchers should determine what type of markers to use based on the following criteria: variability and the number of required markers, the need for their codominance, the corresponding requirements for plant material and isolated DNA, genome size and taxon ploidy; practical - analysis efficiency, cost and related technical support [7]. Moreover, the trend of modern research involving molecular markers is the use of two or more types of markers (for a description of the types, see below), which leads to more unambiguous, reliable results. So, to understand the mechanisms of resistance The V . amurensis to Agrobacterium using RAPD markers in the mapping population of grapes, a resistance locus was identified, on the basis of which resistance-coupled SCAR markers were developed. However, the exact localization of these markers was found out using SSR markers using a reference map [8].

In recent decades, types of molecular markers have been developed based on the hybridization reaction, such as RFLP, and several types, including the PCR stage (RAPD, AFLP, SSR, SNP, CAPS, SCAR, markers based on retrotransposons, etc.). The review [6] provides a list of the most commonly used molecular markers, applications of their use, advantages and disadvantages, as well as a comparative assessment based on a number of features. These include: requirements for the quantity and quality of DNA, the number of analyzed polymorphic loci, ease of use, the ability to automate the process, reproducibility and cost.

CLASSIFICATION OF MOLECULAR MARKERS

Molecular markers are classified into various groups on the basis of:

(1) mode of gene action (co-dominant or dominant markers);

(2) method of detection (hybridization-based molecular markers or polymerase chain reaction (PCR)-based markers);

(3) mode of transmission (paternal organelle inheritance, maternal organelle inheritance, bi-parental nuclear inheritance or maternal nuclear inheritance)

Different types of DNA molecular markers have been developed and successfully applied in genetics and breeding activities in various agricultural crops. The following section provides some brief information related with molecular markers based on their method of detection.

RFLP (restriction fragment length polymorphism; restriction fragment length polymorphism) -the first molecular marker method for DNA profiling. The method is based on the technique of cleavage using special restriction enzymes (restriction endonucleases) of DNA molecules that differ in homologous regions and, accordingly, restriction sites, and comparing the lengths of the obtained fragments of different species and even lines of living organisms. RFLP was the first molecular marker technique and the only marker system based on hybridization. Individuals of same species exhibit polymorphism as a result of insertion/deletions (known as InDels), point mutations, translocations, duplications and inversions. Isolation of pure DNA is the first step in the RFLP methodology. This DNA is mixed with restriction enzymes which are isolated from bacteria and these enzymes are used to cut DNA at particular loci (known as recognition sites). This results in a huge number of fragments with different length. Agarose or polyacrylamide gel electrophoresis (PAGE) is applied for the separation of these fragments by producing a series of bands. Each band represents a fragment having different lengths. Base-pair deletions, mutations, inversions, translocations and transpositions are the main causes for the variation resulting in the RFLP pattern. These variations lead to the gain or loss of recognition sites, resulting in fragments of various length and polymorphism. The restriction enzymes will not cut the fragment if a single base-pair variation occurs in the recognition site. However, if this point mutation occurs in one chromosome but not the other, it is called heterozygous for the marker, as both bands are present [12].The resulting fragments are separated electrophoretic ally. After transfer to the membrane, they are identified by hybridization with radioactively labeled probes (Southern blotting). Hybridization makes it possible to determine the lengths of fragments that are complementary to samples. Each fragment is considered as an allele and is used in genetic analysis. A sufficient degree of polymorphism, codominance, frequency and uniform distribution over the genome, high reproducibility made it possible to use the method in DNA profiling, genome mapping, localization of genes related to diseases, genotyping, the study of phylogenetic. According to PubMed data (plant + RFLP), the maximum number of publications using these markers falls on 2003-2005. Certain disadvantages of the method are the need for large quantities of high quality DNA, the use of isotopes, the duration of the analysis, etc. The corresponding development of more efficient and cheaper types of markers to date has significantly reduced the use of RFLP analysis. The corresponding development of more efficient and cheaper types of tokens has now significantly reduced the use of RFLP analysis. The corresponding development of more efficient and cheaper types of tokens has now significantly reduced the use of RFLP analysis.

POLYMERASE CHAIN REACTION (PCR) MARKERS

PCR (polymerase chain reaction; PCR, polymerase chain reaction) - the use of thermostable DNA polymerase for in vitro amplification of individual / specific DNA sequences or loci using random or specific primers (oligonucleotide sequences). The amplified fragments are separated electrophoretic ally and the bands (peaks) are revealed by staining or autoradiography. The preference for genotyping methods using PCR is primarily due to the need for the reaction of small amounts of DNA (5-100 ng of the sample per reaction).

The discovery and development of the polymerase chain reaction has created opportunities for the design of a wide range of molecular markers. The development of RFLP analysis was the CAPS method(cleaved amplified polymorphic sequence) - PCR amplification of sequences cut out at the restriction site corresponding to the recognition by restriction enzyme of nucleotide sequences of polymorphic homologous regions. The PCR technique was developed by Cary Mullis in 1983, as a technique which could amplify a small quantity of DNA without the application of any living organisms [13]. Denaturation, annealing and extension are the most important steps involved in PCR reactions. For more information about PCR and its protocol, The differences are manifested in products that are easily distinguishable by length - DNA fragments during electrophoresis. CAPS markers are codominant and allow the detection of polymorphism in a large number of individuals. For example, the development of CAPS markers based on the sequences of certain genes and EST sequences allowed the researchers to create the corresponding primer pairs, which were used to identify the authenticity of 67 teas (95%) cultivated in Japan [9]. This type of markers helps to identify recent evolutionary history, to understand the direction of the evolutionary process that determines speciation [10].The development and use of CAPS markers contributes to a more accurate mapping of genes of interest, including resistance to pathogens [11, 12].

RAPD (random amplified polymorphic DNA; amplified DNA fragments) is a method for amplifying DNA segments using random primers (approximately 10 nucleotides) does not require prior knowledge of the DNA sequence, This technique was developed by Williams et al. [17] and Welsh and Mcclelland [18] independently. Amplification of genomic DNA is achieved by PCR using single, short (10 nucleotide) and random primer. During PCR, amplification takes place when two hybridization sites are similar to each other and in opposite direction. These amplified fragments are totally dependent on the length and size of both the target genome and the primer [5]. The selected primer should have minimum 40% GC content, as a primer having less than 40%

GC content will probably not withstand the annealing temperature (72 °C) where DNA elongation occurs by DNA polymerase [17]. For the visualization, the PCR product is then separated in agarose gel stained with ethidium bromide [19]. Polymorphism present either at or between primer binding sites can be detected in the electrophoresis by confirming the presence or absence of specific bands [5]. The quantity and quality of DNA, PCR buffer, magnesium chloride concentration, annealing temperature and Taq DNA (type of DNA polymerase) are some important factors affecting the reproducibility of randomly amplified polymorphic DNA (RAPD) markers [20]. however, due to the stochastic nature of DNA amplification, it is important to optimize and maintain appropriate conditions to obtain reproducible results. The method is used, for example, to study closely related species and molecular identity of plants propagated in vitro.[13, 14, 15]. The disadvantages of the method are: dominance, not very good reproducibility and the need for highly purified non-contaminated DNA, since the use of short random primers can lead to the amplification of fragments of various organisms; locus-non-specificity of markers and non-homology of fragments of the same size (homoplasy). Due to the lack of reproducibility, RAPDs are difficult to use in interlaboratory studies. Due to the above, the significance of the results obtained and their interpretation may be questioned.

Based on the RAPD bands obtained using random primers, their separation, extraction, cloning and sequencing, and the design of specific primers, a type of SCAR markers was developed.(sequence characterized amplified region), the development and use of which has been growing exponentially since the early 1990s. Thus, specific SCAR markers are being developed to discriminate between certain molecular phenotypes of different species of the same genus [16], used in taxonomy [17], discriminating ecotypes [18], identifying unique soma clonal variants and molecular events associated with the variability of soma clones [19]. This type of markers has potential; in breeding [20] it makes it possible to distinguish between certain pathogenic strains of microorganisms [21], etc.

AFLP(amplified fragment length polymorphism; amplification of length polymorphic DNA fragments) - a method of selective PCR amplification of length polymorphic fragments obtained as a result of enzymatic cleavage of genomic DNA with restriction endonucleases. Oligonucleotide adapters are attached to the fragments obtained after restriction for selective amplification. Thus, primers consist of an adapter and a sequence of several nucleotides (1-5) specific for recognition by a restriction enzyme. Fragment polymorphism in length (usually several tens per reaction), caused by multiple genomic restriction sites, is detected by electrophoresis. Reproducibility and a high number of informative fragments in the reaction allows the method to be used in various aspects: the study of genetic identity, identification of varieties and clones, phylogenetic relationships, mapping, MAS, etc. Implementation of the method requires high quality DNA. Although the use of AFLPs, like RAPDs, does not require prior knowledge of DNA sequences, these types of markers are difficult to use when studying different populations or even species. The disadvantages of AFLPs are the dominance of alleles, the possible non-homology of fragments of the same length - homoplasia. In the case of insertion between adjacent restriction sites, the loss of a short fragment and the appearance of a longer one is observed. When using dominant markers, for example, in diploids, one band occurs when one or both homologous chromosomes contain the amplified sequence, i.e. it is impossible to distinguish between homo- and heterozygotes. In polyploidy, the ambiguity is exacerbated, since even if a certain band is identified, it is all the more difficult to say how many alleles are present [22]. However, the use of AFLP markers has remained virtually unchanged over the past decade (PubMed). The method is used to study the evolutionary and ecological aspects of living organisms, while preserving species.

Microsatellites (SSR ,simple sequence repeat; tandems of repeats of simple sequences) - DNA regions consisting of tandems of repeating units: mono-, di-, tri-, tetra- or penta-nucleotides [23]. Microsatellites are present in both non-coding and coding regions of the genome, as well as in the chloroplast [24] and mitochondrial genomes [25]. Natural reasons for the diversity in the number of repetitions of microsatellite units in the genome are the slippage of polymerase during DNA replication and / or inappropriate crossing over, mismatch / repair of DNA double strand damage, and movement of retrotransposons. These variations lead to polymorphism in the length of the fragments detected by electrophoresis [26]. This type of genetic markers has gained great importance in the last decade due to a set of properties: hypervariability,multi-allelic nature, codominant inheritance, high reproducibility, relative abundance, extensive distribution over the genome, high throughput, compliance with process automation. Microsatellites are used in the study of phytogeography, population structure, varietal identification, and identification of the source of origin (paternity).

Compared to SSR, the use of RAPD is not reproducible enough, while RFLP does not have a high throughput, and AFLP is complicated by the fact that individual bands can consist of several fragments of identical size, especially in large genomes. Although for SSRs, homoplasia associated with a high mutation rate and the possibility of reverse mutations can be a problem. One

of the disadvantages of SSR markers is the cost of their development, since the creation of primers requires cloning and sequencing of arbitrary DNA regions, the identification of microsatellite repeats, and the identification of flanking regions of such regions of the genome to determine a specific locus. Microsatellites can be designed from genomic DNA libraries or libraries enriched for specific microsatellites. Some microsatellites can be not just di-, tri-, or tetra-repeats, but compound repeats, in which polymorphism by one nucleotide is possible. These alleles differ by 1 base pair and require special attention during genotyping. Allelic differences in the sizes of microsatellites, on which their use is based, at the same time, can cause errors in amplification and, accordingly, erroneous interpretation of genotyping when assessing genetic diversity, population structure, pedigree, etc. Researchers have shown that the number of errors when using microsatellites directly correlates with the size of the PCR product [27].These alleles differ by 1 base pair and require special attention during genotyping. Allelic differences in the sizes of microsatellites, on which their use is based, at the same time, can cause errors in amplification and, accordingly, erroneous interpretation of genotyping when assessing genetic diversity, population structure, pedigree, etc. Researchers have shown that the number of errors when using microsatellites directly correlates with the size of the PCR product [27].These alleles differ by 1 base pair and require special attention during genotyping. Allelic differences in the sizes of microsatellites, on which their use is based, at the same time, can cause errors in amplification and, accordingly, erroneous interpretation of genotyping when assessing genetic diversity, population structure, pedigree, etc. Researchers have shown that the number of errors when using microsatellites directly correlates with the size of the PCR product [27].that the number of errors when using microsatellites directly correlates with the size of the PCR product [27].that the number of errors when using microsatellites directly correlates with the size of the PCR product [27].

EST - SSR (ESTs ,expressed sequence tag, short fragments of sequences of cloned DNA regions (mRNA → cDNA)), which are used to identify gene transcripts, can be sources for identifying SSRs, i.e. to search for microsatellites in the expressed regions of the genome [28]. Since EST sequences are conserved, interspecies EST-SSRs PCR amplification is arguably more successful than genomic DNA SSRs. This type of markers is of interest because they are inexpensive to develop, they reflect transcribed genes, whose putative functions can often be determined by revealing homology. But the quality of SSR-EST sequences is very important, since the design of primers can be complicated by a number of circumstances: the spread of one or both primers to the splicing sites; the presence of large introns in the genomic DNA sequence, using questionable information regarding primer design, and designing primers for chimeric cDNA clones. Thus, the design of the primers is 60-90% successful. By the beginning of 2013, more than 74 million ESTs are available in the databases (GenBank). However, the creation of gene microsatellite markers is limited to those species for which there are a sufficiently large number of ESTs. Using certain computer programs, sequence data for ESTs, genes and cDNA clones can be downloaded from GenBank and scanned for SSRs [table. 1 in 28]. Applications for the use of this type of markers are functional genome, associative mapping, analysis of genetic diversity, analysis of quantitative traits, etc. EST-SSR markers can contribute to the evolutionary analysis of a wide range of taxa, species analysis, whose resources are extremely limited [29]. The products of EST-SSR markers are distinct bands, distinct allelic peaks [28]. Since these markers are derived from transcripts, they can be used to assess the functional diversity of natural populations or collections of germplasm. Especially in cases of marking genes responsible for phenotypic traits, these markers are useful for use in breeding. As mentioned above, CAPS markers are also being developed based on EST sequences. Responsible for phenotypic traits, these markers are useful in breeding. As mentioned above, CAPS markers are also being developed based on EST sequences. Responsible for phenotypic traits, these markers are useful in breeding. As mentioned above, CAPS markers are also being developed based on EST sequences.

ISSR(inter-simple sequence repeat, genome region between two adjacent, oppositely oriented microsatellites) - the sequence of the core part of the microsatellite with several (1-3) nucleotides adjacent to the tandem of repeats is used as primers. Dozens of fragments of multiple loci obtained by PCR are separated by electrophoresis and evaluated for the presence or absence (due to the dominance of markers) fragments of a certain size. The main advantage of this type of markers is that there is no need to know the sequences when designing primers. It should be borne in mind that due to multilocus, homoplastic is possible. Markers of this type are used to identify genetic identity, pedigree, differentiation of clones, microclines and lines, taxonomy of closely related species.

SNP(single-nucleotide polymorphism; single nucleotide polymorphism) - markers of this type are becoming more and more used in genome research. The technique is based on the fact that in organisms' changes in one nucleotide lead to point mutations, thereby causing polymorphism in one nucleotide (diallelic type of markers). Knowledge of the sequences and flanking regions is required to create specific primers. Despite the high cost, the use of the method has increased in recent years. Since the method allows automated high-resolution genotyping with the simultaneous use of a large number of SNP markers; the presence of many thousands of samples on a chip allows multiple SNPs to be analyzed simultaneously. But at the same time, the difference in alleles

by only one nucleotide and multiple probes makes it impossible to create optimal hybridization conditions for the entire array of probes. In some cases, this leads to hybridization of the analyzed DNA with inappropriate samples. Evaluation of a large amount of data becomes a serious problem, since it is not possible to undertake a point-by-point assessment of the validity of certain data [30].

Since the use of SSR and SNP markers is constantly increasing, researchers have carried out a comparative characterization of these two types of markers [31]. So SSRs have the following advantages over SNPs. At SSR loci, an excess by a certain number of replicates can be considered a polymorphism, while regions of many chromosomes must be sequenced to identify SNPs of homologous regions. SSRs have insignificant bias, i.e. the assessment of polymorphism with their help does not depend on the original list of varieties. The success rate for SSR amplification of closely related cultivars is higher than for SNPs. Loci SSRs are more potent in detecting mixtures than SNPs. The reliability of SSR genotyping is easier to assess, since a larger proportion of errors can be detected in the analysis of the pedigree, when there are many alleles per locus, whereas for balletic SNPs, many errors in the analysis are not detected, since they follow the rules of Mendelian inheritance. Obviously, there are also disadvantages of SSR compared to SNP. Thus, a large number of alleles at the SSR locus implies the analysis of a large number of samples. More frequent spontaneous SSR mutations within a pedigree potentially complicate lineage reconstruction; reverse mutations make it difficult to describe the long history of populations. The variability of highly polymorphic microsatellites may incorrectly reflect genomic diversity. Microsatellites require the inclusion of controls, while the use of SNPs can be performed in parallel in a number of laboratories without the need to calibrate the results. There are also disadvantages of SSR in comparison with SNP. Thus, a large number of alleles at the SSR locus implies the analysis of a large number of samples. More frequent spontaneous SSR mutations within a pedigree potentially complicate lineage reconstruction; reverse mutations make it difficult to describe the long history of populations. The variability of highly polymorphic microsatellites may incorrectly reflect genomic diversity. Microsatellites require the inclusion of controls, while the use of SNPs can be performed in parallel in a number of laboratories without the need to calibrate the results. There are also disadvantages of SSR in comparison with SNP. Thus, a large number of alleles at the SSR locus implies the analysis of a large number of samples. More frequent spontaneous SSR mutations within a pedigree potentially complicate lineage reconstruction; reverse mutations make it difficult to describe the long history of populations. The variability of highly polymorphic microsatellites may incorrectly reflect genomic diversity. Microsatellites require the inclusion of controls, while the use of SNPs can be performed in parallel in a number of laboratories without the need to calibrate the results. Reverse mutations make it difficult to describe the long history of populations. The variability of highly polymorphic microsatellites may incorrectly reflect genomic diversity. Microsatellites require the inclusion of controls, while the use of SNPs can be performed in parallel in a number of laboratories without the need to calibrate the results. Reverse mutations make it difficult to describe the long history of populations. The variability of highly polymorphic microsatellites may incorrectly reflect genomic diversity. Microsatellites require the inclusion of controls, while the use of SNPs can be performed in parallel in a number of laboratories without the need to calibrate the results.

It is obvious that the development and use of various types of molecular markers is based on the identification of changes in genomic sequences, which in turn involves, among other things, extinction (loss) and subsequent insertion of DNA sequences. Retrotransposons (RT) may be the cause of these events. RT are an integral and fairly significant part of the genome of all living organisms. RT-based markers are successfully used to analyze genetic relationships, phylogenetic, species diversity, to create genetic maps and to identify genes [32, 33, 34, 35]. The ubiquitous nature of RT and its involvement in the creation of genomic diversity through the integration of large segments of DNA into dispersed chromosomal loci make them ideal for creating several types of molecular markers on this basis. So,IRAP (inter-retrotransposon amplified polymorphism) generates PCR products between two nearby RTs. REMAP (retrotransposon-microsatellite-amplified polymorphism) uses one LTR (long terminal repeat) primer and another “attached” to the 3 'end of the SSR sequence, revealing PTs integrated around this SSR sequence. iPBS amplification is based on the presence of reverse transcriptase at the primer binding site in the LTR region of PT. SSAP (sequence-specific amplified polymorphism) is an AFLP derivative that amplifies products between the PT integration site and the restriction site to which the adapter is bound.

Types, Causes And Consequences Of Genotyping Errors

Obviously, the reliability of conclusions for any genotyping experiment is determined by the quality of the results obtained. Genotyping errors can be due to various reasons. Their classification is given in reviews [27, 36, 37, 38]. The minimization of errors is possible by assessing the characteristics of the experimental methods used, their optimization, the proper use of controls,

replicates, the selection of markers, as well as the development of statistical approaches to identify errors. Thus, an increase in the probability of errors is associated with markers with higher heterozygosity, with a large number of alleles, a large number of "stutter" bands, and large product sizes [27]. The presence of statter bands makes it difficult to assess heterozygosity, for example, when two alleles of a heterozygote differ by only one repeat, stutters overlap them, forming adjacent bands of similar intensity.

Four types of genotyping errors tend to increase the proportion of homozygosity: loss of alleles, null alleles, incorrect assessment of nearby alleles as “stutter” bands, dominance of short alleles, when larger alleles have a weak signal below the detection threshold [38]. Genotyping errors can manifest themselves in a violation of Mendelian inheritance, but there are those in the presence of which the result is still in accordance with Mendelian inheritance. While the latter are more difficult to identify, they can have a significant impact on the validity of the statistical analysis. Even such errors identified can often be attributed to more than one source within the pedigree [39]. So SNPs are balletic markers, and their inheritance corresponds to Mendelian [40]; nevertheless,this type of markers can cause a greater proportion of errors. Moreover, more and more data sets contain a number of errors due to researcher oversights, software deficiencies, or simply biochemical anomalies, which further exacerbate the use of such high-performance methods as SNP [41].

Genotyping errors can lead to incorrect genetic maps when studying the mapping of signs of interest, incorrect determination of the pedigree, their influence on the genetic analysis of populations is obvious. The importance of detecting genotyping errors becomes more and more obvious with the development of international programs, for example, for the study of biodiversity and the need to process huge amounts of data obtained in a number of laboratories, when the reliability and reproducibility of the results become of decisive importance.

Errors are divided by researchers into categories [37].

1 - due to the peculiarities of the actual sequence of the DNA molecule. In the case of mutations localized in the complementary sequence of one of the markers, amplification is impaired; an insertion or deletion close to a microsatellite may result in the same fragment size or "read" only one allele instead of two.

2 - low quality and / or insufficient amount of DNA. Both low quality and small amount of DNA cause "dropout" of alleles in heterozygotes (amplification of only one, preferably shorter allele). If DNA samples are contaminated, the amplification of the contaminant alleles is possible. One of the solutions to the problem in the case of a small amount of DNA was proposed a method for dividing a sample into several tubes and carrying out an amplification reaction in each of them [42].

3 - artifacts caused by reagents. Errors associated with the feature of Tag polymerase to add a nucleotide that does not belong to the sequence of the amplified fragment (usually adenine) to the 3′-end; as a result, an additional band (peak) may appear. This reaction is sensitive to the sequence of the 5 'end of the primer and the long elongation times in PCR. Slippage of Tag polymerase in the early stages of PCR leads to the appearance of false alleles. Poor quality reagents can adversely affect the PCR conditions, inadequate electrophoresis conditions can lead to a distorted pattern of allele size distribution, etc.

4 - human factor. A significant proportion of genotyping errors is due to this very factor [36]. So, this may be due to the subjectivity of the assessment of electrophoregrams and radio autograms, and the corresponding incorrect designation of alleles, which in turn depends on the quality of the data; contamination of exogenous DNA or accidental cross-contamination of samples, use of suboptimal protocols, suboptimal primers, PCR melting points, errors in data entry and processing, etc. [43]. In practice, to detect genotyping errors, a repeated amplification reaction and a comparison of the obtained products of the initial and repeated analysis for a certain number of samples are carried out, in other words, the analysis of duplicated samples.

The number (coefficient, percentage) of genotyping errors is estimated in various experiments per allele, per locus, PCR, per multilocus. However, since the errors are not randomly distributed among alleles, loci, PCR, a simple comparison of the percentage of errors between them is incorrect [38]. So, when comparing the percentage of error per locus, the loss of the allele in the homozygous locus does not appear, and the loss of the allele in the heterozygous one will look like a homozygote. Therefore, allelic loss in a homozygote and heterollocus with a lost allele will be estimated equally and, therefore, the percentage of error between loci or populations that actually differ in the degree of heterozygosity is estimated incorrectly. Determination of the fraction of error per locus is considered the most acceptable. The percentage of error for a particular allele or locus gives certain information on their expediency of "excluding" from the experiment to increase the reliability of the results. What percentage of the error can be significant for the correctness of the assessment of the results can be seen from the example: with a 1% error in the designation of

the allele, only in 62% of cases repeated comparison of the individual's genotype confirmed the identity, with 2% the indicator decreased to 40% [27]. In simulated studies, it was demonstrated that an error of 3% negatively affected the identification of linkage disequilibrium (LD), which interferes, for example, with the identification of relationships in the complexes of genes responsible for the disease [44]. with 1% error in the designation of the allele, only in 62% of cases repeated comparison of the genotype of an individual confirmed the identity, with 2% the indicator decreased to 40% [27]. In simulated studies, it was demonstrated that an error of 3% negatively affected the identification of linkage disequilibrium (LD), which interferes, for example, with the identification of relationships in the complexes of genes responsible for the disease [44]. with 1% error in the designation of the allele, only in 62% of cases repeated comparison of the genotype of an individual confirmed the identity, with 2% the indicator decreased to 40% [27]. In simulated studies, it was demonstrated that an error of 3% negatively affected the identification of linkage disequilibrium (LD), which interferes, for example, with the identification of relationships in the complexes of genes responsible for the disease [44]. responsible for the disease [44]. responsible for the disease [44].

According to the definition of researchers [36], it should be remembered that the potential consequences of errors on conclusions are inversely proportional to the scale of the conclusions. There is no universal approach for eliminating genotyping errors, at least for a number of objective reasons for their nature. Nevertheless, researchers [37] formulate defining attitudes to reduce the number of errors and reduce their impact on the final result. Among them - an assessment of the quality of samples intended for analysis (DNA quality) and the level of technical capabilities for implementation; conducting pilot experiments for a comparative assessment of the theoretical and real error rate; quality control should be carried out at all stages of the study to identify possible large variants of errors. Saving material and labor costs should be subordinated to reducing the number of errors and increasing the efficiency of the analysis of the results obtained. Only highly qualified personnel, guided and using the quality standards of testing procedures, with the least possible arbitrary manipulations, with the greatest possible automation of processes, is able to minimize the percentage of genotyping errors.

A number of approaches have been developed to identify genotyping errors and determine their percentage, which include comparison of duplicated samples, independent designation of allele sizes, and verification of compliance with Mendelian inheritance [45]. But none of the approaches is capable of identifying all errors. The researchers compared a commercial genome wide set and custom-designed fine-resolution mapping set. The creation of a commercial genome wide set is based on both localization and the ability to amplify DNA under normal conditions. Therefore, all problematic primers are replaced with optimal ones. When designing primers for the fine-resolution mapping set, markers are selected according to their chromosomal localization, and their amplifying ability is not always optimal. This feature was reflected in the initial PCR failures. Thus, the errors of Mendelian inheritance were 0.13% for the first set of primers and 1.19% for the second because of the difficulties in calculating some alleles in the second case. Concordance checking revealed only human errors: missing alleles, misidentified and confused samples, while Mendelian inheritance showed additional errors such as mutations and null alleles. The latter approach was more reliable when using both types of microsatellite sets, while the dedicated mapping marker set was more successful when comparing duplicate samples.19% for the second because of the difficulty of counting some alleles in the second case. Concordance checking revealed only human errors: missing alleles, misidentified and confused samples, while Mendelian inheritance showed additional errors such as mutations and null alleles. The latter approach was more reliable when using both types of microsatellite sets, while the dedicated mapping marker set was more successful when comparing duplicate samples.19% for the second because of the difficulty of counting some alleles in the second case. Concordance checking revealed only human errors: missing alleles, misidentified and confused samples, while Mendelian inheritance showed additional errors such as mutations and null alleles. The latter approach was more reliable when using both types of microsatellite sets, while the dedicated mapping marker set was more successful when comparing duplicate samples. Whereas Mendelian inheritance showed additional errors such as mutations and null alleles. The latter approach was more reliable when using both types of microsatellite sets, while the dedicated mapping marker set was more successful when comparing duplicate samples. Whereas Mendelian inheritance showed additional errors such as mutations and null alleles. The latter approach was more reliable when using both types of microsatellite sets, while the dedicated mapping marker set was more successful when comparing duplicate samples.

The most obvious of the approaches - repeated genotyping and comparison of a representative number of samples, although it is associated with rather large additional experimental costs, still does not provide a complete guarantee of a reliable result. The most commonly used statistical test in the analysis of populations is the determination of compliance / deviation with the Hardy – Weinberg equilibrium (HWE) (Deviations from Hardy – Weinberg equilibrium, HWD) [46]. The Hardy-Weinberg principle

describes how variation in a population is maintained based on Mendelian inheritance. The Hardy-Weinberg law states that the frequency of alleles (i.e., gene variations) and genotype in a population, due to Mendelian inheritance, in the absence of evolutionary influences, remains constant across generations. A number of assumptions are required to comply with this principle: diploid organisms, only sexual reproduction, generations do not overlap, do not duplicate, random crosses (with inbreeding, the homozygosity of all genes increases), an indefinitely large population size, there are no migrations, mutations and selection.

At a locus with two alleles (p + q) = (p + q) 2 = p 2 + 2 pq + q 2 = 1, i.e. the proportions of genotypes, according to Hardy-Weinberg, are p 2 +2 pq + q 2 (binomial distribution).

For more than two alleles (p + q + r) 2 = p 2 + q 2 + r 2 +2 pq +2 pr +2 qr .

In a more general form for alleles A 1 , ... A n with the frequency of alleles from p 1 to p n (p 1 + ... + p n) 2 .

Deviations from the Hardy-Weinberg equilibrium may indicate genotyping errors. But it should be remembered that deviations can be caused by the very structure of the population. What, in fact, is important to identify and differentiate.

To assess the extent to which the analyzed data coincide with the theoretically expected, Pearson's “test of fit”, the chi-square method (χ 2), is used. Two sets are compared: one is an empirical frequency distribution, the other is a sample with the same parameters (n , M , S , etc.) as the empirical one, but its frequency distribution is constructed in exact accordance with the chosen theoretical law, which is supposedly subject to behavior of the studied random variable. In general terms, the formula for the compliance criterion can be written as follows:

Figfor1

where α is the actual frequency of observations;

A is the theoretical expected frequency for this class.

Unfortunately, checking for inconsistency with Mendelian inheritance does not exclude all genotyping errors [40]. The researchers analyzed 107,000 generated genotypes using various approaches, including TaqMan, RFLP, sequencing, mass spectroscopy, SNPs (distributed in both gene and intergenic regions). Of 313 SNPs, the frequencies of minor alleles of which were in the range of 0.06–0.49, in 36 cases (11.5%) significant deviations from HWE were observed. Genotyping errors were found for 21 SNPs; for 5 SNPs - non-specific binding; for 10 SNPs, the reasons for the deviation were not clarified. According to each of the technologies used and the level of significance of deviations from HWE (0.01 < P <0.05 or P<0.01), it was concluded that the SNP methodology was at least partly the source of the error. The result of a retrospective analysis of these experiments to identify the proportion of nonspecific samples and genotypic errors allowed the authors to develop and standardize the process and, in further studies on genotyping 96 samples using 1434 SNPs, to reduce the HWE to 10 (P <0.01).

It can be assumed that an insignificant percentage of errors in genotyping populations using microsatellites with a significant number of samples and a large number of alleles will not have a serious impact on the characteristics of the population structure. However, it was found that errors due to the characteristics of the samples can significantly affect the result, despite the large sample [38]. To assess the effect of individual samples on the Hardy-Weinberg equilibrium, a jackknife analysis was used, which was carried out using appropriate programs. Each sample was sequentially removed from the general data set, and the Hardy-Weinberg index for all alleles was calculated for the remaining samples. "Influential" samples were considered those, at removal of which the level of significance became P> 0.05, while for the entire sum of samples it was P ≤ 0.05. There were 33 cases where the removal of the sample changed the status of the locus as influencing the deviation of equilibrium to being in equilibrium (P > 0.05).

The magnitude of the influence was determined by transforming the value of P:

figfor2

Thus, it was shown that the Hardy-Weinberg equilibrium was very sensitive to homozygosity of rare alleles in individual individuals, and more than 50% of these cases were due to genotyping errors of poor quality samples. According to the researchers' definition, this indicates the possibility that even a “normal” level of laboratory errors may cause an overestimation of the effect of individual markers on the Hardy-Weinberg equilibrium and thus on the assessment of population structure, including the “overestimation” of the role of inbreeding. Any error that significantly affects the distribution frequency contributes to an incorrect reflection of the degree of population differentiation. For example, if a larger allele occurs in only one of two populations, a dropout

error makes the two formations more similar. On the contrary, if the same allele is present in two populations, its loss, reducing its frequency, leads to the fact that the significance of differentiation is determined by secondary differences in the frequencies of rare alleles.

Back in the early 90s, when studying the human genome, data appeared on the non-amplification of individual alleles, for example, 7 out of 23 alleles (30%) on chromosome 16 were not amplified in some representatives of the pedigree [47]. When genotyping Cervus elaphus3 alleles out of 16 were not amplified, which was revealed by the mismatch in mother-offspring pairs and a significant deviation from the Hardy-Weinberg equilibrium [48]. Analysis of two out of three alleles showed that the exclusion of even a few unamplified homozygotes can have a serious effect on the interpretation of the frequency distribution in the genotype and the incorrect assessment of inbreeding in the population. Since many natural populations exhibit Hardy-Weinberg disequilibrium as a consequence of inbreeding or mixing, it is necessary to differentiate between these cases and deviations from equilibrium due to zero alleles. Researchers have proposed a calculation algorithm that can be used to separate loci exhibiting zero alleles from those reflecting the biological signal of inbreeding [49].

At present, the phenomenon of non-amplification of microsatellite alleles has already been unambiguously recognized. Null alleles are detected not only with the use of microsatellites, but also with EST-SSR [28]. The molecular nature of null alleles is due to mutations: substitutions, insertions, deletions. The relationship between the presence of zero alleles and the high variability of the flanking regions has been recorded in a number of molecular genetic studies. Thus, computer simulation carried out by researchers and its application to empirical data allowed the conclusion that null alleles were observed in populations with a high level of diversity in the flanking areas (≥0.001) [50]. Null alleles with a high probability were fixed in populations of large effective size with a high proportion of mutations in the flanking regions and this differed from the population ("central"),on the basis of which the alleles were cloned and primers were designed. In other words, the high frequency of null alleles in the studied populations was due to the high level of differentiation between them and the "central" population.

Errors due to "stuttering" (due to the characteristics of Tag polymerase), the loss of larger alleles and zero alleles, in contrast to stochastic (random, random) errors, create consistent deviations from the actual pattern of alleles and, accordingly, deviations from the adequacy of genotyping [51]. Thus, the size and shape of stuttering samples vary among loci, some markers exhibit a low level of stuttering, while others can form two or more peaks. Interpretation of such samples is difficult, for example, adjacent alleles in a heterozygote differing by one dinucleotide repeat. Such an allele of a heterozygote can be estimated as a homozygote with a larger allele. In general, this will lead to a shift towards larger alleles, reducing heterozygosity and increasing the apparent level of inbreeding of the corresponding loci. With regard to the loss of a larger allele, due to the predominant amplification of the shorter allele, the peak of the latter may be much higher than the first, and in low quality samples, the larger allele is not amplified at all. The loss of alleles increases with their size, and in some loci, even an increase in DNA concentration does not decrease the level of loss, indicating technical limitations in the amplification of large alleles. Null alleles - the third reason for consecutive genotyping errors, are not detected by definition, the reason for them is mainly a mutation at the priming site (primer attachment). In the case of a heterozygote, the samples will be judged as homozygote, and in the case of a homozygote, as no amplification reaction at all. Generally, this leads to a decrease in the frequency of heterozygosity and an increase in the apparent level of inbreeding. Researchers recommend standard procedures and protocols to prevent and reduce genotyping errors when studying wild populations [52]. These include screening (selection) of microsatellite loci, reanalyzing a subset of samples, evaluating the complete dataset, testing errors, reducing errors in subsequent analyzes, and reporting error rates. Testing errors, reducing errors in subsequent analyzes, and reporting on error rates. Testing errors, reducing errors in subsequent analyzes, and reporting on error rates.

In [53], when analyzing 233 publications with mentioned zero alleles, attention was drawn to several provisions: how these alleles were determined; whether primers for these loci have been redesigned; what method was used to determine their frequency; whether they were sequenced to elucidate the molecular basis of their "manifestation"; whether loci with null alleles were retained or excluded from subsequent analysis. The approach most often used in these works for identifying zero alleles was based on the observed deficiency of heterozygotes in populations. An estimate of the deviations from the Hardy-Weinberg equilibrium was also used. Although this indicator is also influenced by the structure of subpopulations, inbreeding or selection, which affected a place close to the microsatellite. In several works, the analysis of the pedigree was involved, which is more significant evidence. The authors concluded that the frequencies of null alleles could lead to false conclusions about pedigree. Zero alleles and the observed high homozygosity can lead to an overestimation of population differentiation. With a high differentiation of populations, the

presence of zero alleles leads to an overestimation of the fixation index (F ST) - measures of population differentiation and genetic distances.

Since the primers flitting EST-SSRs are derived from fairly conserved sequences, it is likely that null alleles may be less of a problem for these markers. Comparison of 25 EST-SSRs and 17 SSRs markers in genotyping of a number of conifers (Picea spp .) Showed less heterozygosity for EST-SSRs markers (mean H = 0.65 vs 0.72). But at the same time, the values of the probability of inbreeding (F) for specific SSR markers made it possible to identify loci with imaginary null alleles, as a result, indicating a significantly lower apparent inbreeding (mean F = 0.046 vs 0.126) [54].

As mentioned above, for molecular markers, homoplasia is possible - the non-homology of amplified sequences of the same size. The use of microsatellites and sequencing of DNA fragments has significantly increased the ability of researchers to identify signs of a homoplastic nature and their effect on population analysis [55]. The original concept of homoplasia was used by evolutionists to describe the fact that a trait present in two species may not have one common ancestor, but may be the result of convergence, parallelism, or reversion. Thus, homoplasia is important in assessing phylogenetic relationships between species. For a molecular marker, homoplasia occurs when different copies of a locus are the same size and are not electrophoretically separated (electromorphs),but not identical in origin. For example, microsatellites mutate at a fairly high rate and alleles of the same size have hidden sequence differences that can only be detected by sequencing or (SSCP, single-strand conformation polymorphism) [56]. Homoplasia in size within and between populations prompts researchers to define its index as the probability that at a given locus two copies of a gene of the same condition are not identical in origin.Homoplasia in size within and between populations prompts researchers to define its index as the probability that at a given locus two copies of a gene of the same condition are not identical in origin.Homoplasia in size within and between populations prompts researchers to define its index as the probability that at a given locus two copies of a gene of the same condition are not identical in origin.

Even more significant for the interpretation of the results is the manifestation of homoplasia in AFLP analysis. Thus, since the proportion of AFLP profiles of the same size ranged from 0 to 5% for close relatives and increased manifold for distant taxa, researchers suggested that the study of phylogeny using AFLPs is possible only for closely related taxa [57]. Using AFLP in the comparative characterization of Phaseolus lunatus and Lolium perenne ,researchers have shown that for both species the fragment size distribution is asymmetric, with a much higher proportion of short fragments compared to long ones [58]. Smaller fragments are more susceptible to combination during electrophoresis and manifestation of homoplasticity. The researchers hypothesized that the level of homoplastic in size is a function of size and that this may bias estimates of genetic diversity based on the frequencies of AFLP fragments. In silico analysis of the complete genomic sequences of 14 species, including eukaryotes, prokaryotes, and archaea, the researchers revealed the following: the previously shown positive correlation between the percentage of homoplastic and genome size is a direct consequence of the number of bands observed and the content of GC bases in the genome [59].The regression of the proportion of homoplastic by the number of bands was identical for all analyzed species, despite the 1,000-fold difference in genome size.

Since researchers admit that genotyping using molecular markers is a highly subjective procedure depending on the training of personnel and the capabilities of a particular laboratory, as one of the criteria, for example, for AFLP markers, the scan AFLP marker selection algorithm was developed based on the intensity of distribution of fragments among samples [60]. The method is compared with the previously proposed AFLP score [56], in which the thresholds of the fluorescence intensity of the AFLP polymerase reaction product were used as a criterion. Algorithms were used to scan a population of 619 Arabas alpine individuals of 191 collection sites and assessment of genetic structure, diversity and differentiation. In each 96-well plate, the following genotyping controls were used, starting with DNA isolation: one negative control (no sample), two samples included in all plates (replicates, positive control for all plates) and two samples per plate were analyzed twice (positive control on the die). In addition, one sample was collected as a duplicate at each collection site. Of these, 39 randomly selected were included in the analysis as blind controls to calculate the discrepancy in the error rate. The automated marker selection procedure has significantly reduced the error rate by lowering the random signals associated with both individuals and populations. The randomly selected marker sets showed markedly less variability between duplicates compared to the sets selected according to a specific criterion. Attention should be paid to the set of controls for genotyping. A similar approach is required when using any type of molecular marker.

Obviously, along with the need to identify and reduce the proportion of genotyping errors that occur in any experiment when using different types of markers, it is important to maintain a trade-off between culling error-generating loci and increasing the potential of the retained loci to enhance the genetic signal of the population. The compromise can be different in different studies,

but the main thing is that this signal is not lost for the sake of an “acceptable” level of error, and the researchers propose an empirical approach to achieve the desired compromise [61].

CONCLUSION

Over the past two decades, there has been an exponential growth in the use of molecular markers in the study of various aspects of the vital activity of plant organisms. The total number of works is already approaching two tens of thousands. In this regard, it is obvious not only the importance of using different types of molecular markers, but also the correctness of the assessment of the results obtained with their help. Based on the vast experimental experience, in a number of reviews, researchers formulated the advantages and disadvantages of the used molecular markers, the main requirements for their selection for one or another kind of research. Despite the variety of markers, there are a number of primary general requirements and criteria to improve the reliability of the results. Among them, the choice of an adequate type of markers for solving the task; sufficiently high polymorphism of markers, high resolution, codominance, frequency and uniformity of distribution over the genome, high reproducibility. No less important and essential in genetic analysis is the minimization of errors, which are divided by researchers into categories. The main ones are: the appropriate quality and quantity of analyzed DNA; assessment and selection of optimal primers to increase the information content of the corresponding loci; features of Tag polymerase and optimization of PCR conditions; the proper use of replicates and controls; and the adequacy of statistical approaches for evaluating the results obtained and detecting errors in analysis. No less important and essential in genetic analysis is the minimization of errors, which are divided by researchers into categories. The main ones are: the appropriate quality and quantity of analyzed DNA; assessment and selection of optimal primers to increase the information content of the corresponding loci; features of Tag polymerase and optimization of PCR conditions; the proper use of replicates and controls; and the adequacy of statistical approaches for evaluating the results obtained and detecting errors in analysis. No less important and essential in genetic analysis is the minimization of errors, which are divided by researchers into categories. The main ones are: the appropriate quality and quantity of analyzed DNA; assessment and selection of optimal primers to increase the information content of the corresponding loci; features of Tag polymerase and optimization of PCR conditions; the proper use of replicates and controls; and the adequacy of statistical approaches for evaluating the results obtained and detecting errors in analysis. Features of Tag polymerase and optimization of PCR conditions; the proper use of replicates and controls; and the adequacy of statistical approaches for evaluating the results obtained and detecting errors in analysis. Features of Tag polymerase and optimization of PCR conditions; the proper use of replicates and controls; and the adequacy of statistical approaches for evaluating the results obtained and detecting errors in analysis.

REFERENCES

[1] Ganal MW, Polley A., Graner EM, Plieske J., Wieseke R., Luerssen H., Durstewitz G. Large SNP arrays for genotyping in crop plants // J Biosci. - 2012. - Vol. 37. - P. 821-288.

[2] 2.Miedaner T., Korzun V. Marker-assisted selection for disease resistance in wheat and barley breeding // Phytopathology. - 2012. - Vol.102. - P. 560-566. Paux E., Sourdille P., Mackay I.,

[3] Feuillet C. Sequence-based marker development in wheat: advances and applications to breeding // Biotechnol Adv. - 2001. - Vol. 30. - P. 1071-1088.

[4] Hall D., Tegström C., Ingvarsson PK Using association mapping to dissect the genetic basis of complex traits in plants // Brief Funct Genomics. - 2010. - Vol. 9. - P. 157-165.

[5] Zimmer EA, Wen J. Using nuclear gene data for plant phylogenetics: progress and prospects // Mol Phylogenet Evol. - 2012. - Vol. 65. - P. 774-785.

[6] Kumar P., Gupta VK, Misra AK, Modi DR, Pandey BK Potential of Molecular Markers in Plant Biotechnology // Plant Omics Journal. - 2009. - Vol. 2. - P. 141-162.

[7] Woodhead M., Russell J., Squirrell J., Hollingsworth PM, Mackenzie K., Gibb M., Powell W. Comparative analysis of population genetic structure in Athyrium distentifolium (Pteridophyta) using AFLPs and SSRs from anonymous and transcribed gene regions // Molecular Ecology. - 2005. - Vol. 14. - R. 1681-1695.

[8] Kuczmog A., Galambos A., Horváth S., Mátai A., Kozma P., Szegedi E., Putnoky P. Mapping of crown gall resistance locus Rcg1 in grapevine // Theor Appl Genet. - 2012. - Vol.125. - P. 1565-1574.

[9] Ujihara T, Taniguchi F, Tanaka J, Hayashi N. Development of Expressed Sequence Tag (EST) -based Cleaved Amplified Polymorphic Sequence (CAPS) markers of tea plant and their application to cultivar identification // J Agric Food Chem. - 2011. - Vol. 59. - P. 1557-1564.

[10] Segatto ALA, Caze ALR, Turchetto C., Klahre U., Kuhlemeier C., Bonatto SL, Freitas LB Nuclear and plastid markers reveal the persistence of genetic identity: A newperspective on the evolutionary history of Petunia exserta // Molecular Phylogenetics and Evolution. - 2014. - Vol. 70. - P. 504-512.

[11] Cui Y., Lee MY, Huo N., Bragg J., Yan L., Yuan C., Li C., Holditch SJ, Xie J., Luo MC, Li D., Yu J., Martin J., Schackwitz W., Gu YQ, Vogel JP, Jackson AO, Liu Z., Garvin DF Fine mapping of the Bsr1 barley stripe mosaic virus resistance gene in the model grass Brachypodium distachyon // PLoS One. - 2012. - Vol. 7: e38333.

[12] Jo K.-R., Arens M., Kim T.-Y., Jongsma MA, Visser RGF, Jacobsen E., Vossen HJ Mapping of the S. demissum late blight resistance gene R8 to a new locus on chromosome IX // Theor Appl Genet. - 2011. - Vol. 123. - P. 1331-1340.

[13] Shu QY, Liu GS, Qi DM, Chu CC, Liu. J, Li HJ An effective method for axillary bud culture and RAPD analysis of cloned plants in tetraploid black locust // Plant Cell Rep. - 2003. - Vol. 22 .-- P. 1751-1780.

[14] Sreedhar RV, Venkatachalam L., Bhagyalakshmi N. Genetic fidelity of long-term micropropagated shoot cultures of vanilla (Vanilla planifolia Andrews) as assessed by molecular markers // Biotechnol. J. - 2007. - Vol. 2. - P. 1007-1013.

[15] Rai GK, Singh M., Rai NP, Bhardwaj DR, Kumar S. In vitro propagation of spine gourd (Momordica dioica Roxb.) And assessment of genetic fidelity of micropropagated plants using RAPD analysis // Physiol Mol Biol Plants. - 2012. - Vol. 18. - P. 273-280.

[16] Kim JK, An GH, Ahn SH, Moon YH, Cha YL, Bark ST, Choi YH, Suh SJ, Seo SG, Kim SH, Koo BC Development of SCAR marker for simultaneous identification of Miscanthus sacchariflorus, M. sinensis and M. x giganteus // Bioprocess Biosyst Eng. - 2012. - Vol.35. - P. 55-59.

[17] Pankin AA, Khavkin EE Genome-specific SCAR markers help solve taxonomy issues: a case study with Sinapis arvensis (Brassiceae, Brassicaceae) // Am J Bot. - 2011. - Vol. 98: e54-7.

[18] Ray T., Roy SC Genetic diversity of Amaranthus species from the Indo-Gangetic plains revealed by RAPD analysis leading to the development of ecotype-specific SCAR marker // J Hered. - 2009. - Vol. 100. - P. 338-347.

[19] Osipova ES, Lysenko EA, Troitsky AV, Dolgikh YI, Shamina ZB, Gostimskii SA Analysis of SCAR marker nucleotide sequences in maize (Zea mays L.) somaclones // Plant Sci. - 2011. - Vol. 180. - P. 313-322.

[20] Rahman M., Sun Z., McVetty PB, Li G. High throughput genome-specific and gene-specific molecular markers for erucic acid genes in Brassica napus (L.) for marker-assisted selection in plant breeding // Theor Appl Genet. - 2008. - Vol. 117. - P. 895-904.

[21] Naeimi S., Kocsubé S., Antal Z., Okhovvat SM, Javan-Nikkhah M., Vágvölgyi C., Kredics L. Strain-specific SCAR markers for the detection of Trichoderma harzianum AS12-2, a biological control agent against Rhizoctonia solani , the causal agent of rice sheath blight // Acta Biol Hung. - 2011, Mar. - No. 62 (1). - R. 73-84.

[22] Falush D., Stephens M., Pritchard JK Inference of population structure using multilocus genotype data: dominant markers and null alleles // Molecular Ecology Notes. - 2007. - Vol. 7. - P. 574-578.

[23] Powel W., Machray GC, Provan J. Polymorphism revealed by simple sequence repeats // Trends Plant Sci. - 1996. - Vol. 1. - P. 215-222.

[24] Chung AM, Staub JE, Chen JF Molecular phylogeny of Cucumis species as revealed by consensus chloroplast SSR marker length and sequence variation // Genome. - 2006. - Vol. 49.

[25] Rajendrakumar P., Biswal AK, Balachandran SM, Srinivasarao K., Sundaram RM Simple sequence repeats in organellar genomes of rice: frequency and distribution in genic and intergenic regions // Bioinformatics. - 2007. - Vol. 23. - P.1-4.

[26] Kalia RK Rai MK, Kalia S., Singh R., Dhawan AK Microsatellite markers: an overview of the recent progress in plants // Euphytica. - 2011. - Vol.177. - P.309–334

[27] Hoffman JI, and Amos W. Microsatellite genotyping errors: detection approaches, common sources and consequences for paternal exclusion // Molecular Ecology. - 2005. - Vol. 14. - P.599-612.

[28] Varshney RK, Graner A., Sorrells ME Genic microsatellite markers in plants: features and applications // TRENDS in Biotechnology. - 2005. - Vol. 23. - P.48-55.

[29] Ellis R., JM Burke. EST-SSRs as a resource for population genetic analyzes // Heredity. - 2007. - Vol. 99. - P.125-132.

[30] Saunders IW, Brohede J., Hannan GN Estimating genotyping error rates from Mendelian errors in SNP array genotypes and their impact on inference // Genomics. - 2007. –Vol. 90. - P. 291-296.

[31] Guichoux E., Lagache L., Wagner S., Chaumeil P., Le. Ger P., Lepais O., Lepoittevin C., Malausa E., Revardel E., Salin F., Petit RJ Current trends in microsatellite genotyping // Molecular Ecology Resources. - 2011. - Vol. 11. - P. 591-611.

[32] Schulman AH, Flavell AJ, Ellis TH The application of LTR retrotransposons as molecular markers in plants // Methods in Molecular Biology. - 2004. - Vol. 260. - R.145-173.

[33] Kalendar R., Antonius K., Smykal P., Schulman AH iPBS: a universal method for DNA Wngerprinting and retrotransposon isolation // Theor Appl Genet. - 2010. –121. - R. 1419-1430.

[34] Castro I., D'Onofrio C., Martin JP, Ortiz JM, De Lorenzis G., Ferreira VO Pinto-Carnide. Effectiveness of AFLPs and Retrotransposon-Based Markers for the Identification of Portuguese Grapevine Cultivars and Clones Mol Biotechnol. - 2012. - Vol. 52. - P. 26–39.

[35] Nakatsuka T., Yamada E., Saito M., Hikage T., Ushiku Y., Nishihara M. Construction of the first genetic linkage map of Japanese gentian (Gentianaceae) // BMC Genomics. - 2012. –Vol. 13. - P. 672.

[36] Bonin A., Bellemain E., Bronken Eidesen P., Pompanon F., Brochmann C., Taberlet P. How to track and assess genotyping errors in population genetics studies // Mol Ecol. - 2004. - Vol. 13. - P.3261-3273.

[37] Pompanon F., Bonin A., Bellemain E., Pierre Taberlet. Genotyping Errors: Causes, Consequences And Solutions. Genetics. - 2005. - Vol. 6. - P.847-859.

[38] Morin PA, Leduc RG, Archer FI, Martien KK, Huebinger R., Bickham JW, Taylor BL Significant deviations from Hardy – Weinberg equilibrium caused by low levels of microsatellite genotyping errors // Molecular Ecology Resources. - 2009. - Vol. 9. - P.498-504.

[39] Gordon D., Heath SC, Ott J. True pedigree errors more frequent than apparent errors for single nucleotide polymorphisms // Hum Hered. - 1999. - Vol.49. - P.65–70.

[40] Gordon D., Finch SJ, Nothnagel M., Ott J. Power and sample size calculations for case – control genetic association tests when errors are present: application to single nucleotide polymorphisms // Human Heredity. - 2002. - Vol. 54. - P.22-33.

[41] Sobel E., Papp JC, Lange K. Detection and integration of genotyping errors in statistical genetics Am. J. Hum. Genet. - 2002. - Vol. 270. - P.496-508.

[42] Navidi W., Arnheim N., Waterman MS A multiple-Tubes Approach for Accurate genotyping of very Small DNA samples by using PCR: Statistical considerations // Am. J. Hum. Genet. - 1992. - Vol. 50. - P.347-359.

[43] Ghosh, S. et al. Methods for precise sizing, automated binning of alleles, and reduction of error rates in large-scale genotyping using fluorescently labeled dinucleotide markers // Genome Res. - 1997. - Vol. 7. - P. 165-178.

[44] Hosking L., Lumsden S., Lewis K., Yeo A., McCarthy L., Bansal A., Riley J., Purvis I., a Xu Ch-F. Detection of genotyping errors by Hardy – Weinberg equilibrium testing // European Journal of Human Genetics. - 2004. - Vol. 12. - P. 395-399.

[45] Ewen KR, Bahlo M., Treloar SA, Levinson DF, Mowry B., Barlow JW, SJ Foote. Identification and Analysis of Error Types in High-Throughput Genotyping // Am. J. Hum. Genet. - 2000. - Vol. 67. - P.727-736.

[46] Gomes I., Collins A., Lonjou C. et al. Hardy – Weinberg quality control. Annals of Human Genetics. - 1999. - Vol. 63. - P. 535-538.

[47] Callen DF, Thompson AD, Shen Y., Phillips H., Richards Rl., Mulley JC, Sutherland GR Incidence and origin of 'null' alleles in the (AC) n microsatellite markers // Americun Jounrnl of Human genetics. - 1993. - Vol. 52. - P.922-927.

[48] Pemberton JM, Slate J., Bancroftt 'DR, Barrett JA Nonamplifying alleles at microsatellite a caution for parentage and population loci: studies // Molecular Ecology. - 1995. - Vol. 4. - P.249-252.

[49] Van Oosterhout C., Weetman D., Hutchinson WF Estimation and adjustment of microsatellite null alleles in nonequilibrium populations // Molecular Ecology Notes. - 2006. - Vol. 6. - P.255-256.

[50] Chapuis M.-P., Estoup A. Microsatellite Null Alleles and Estimation of Population Differentiation // Mol. Biol. Evol. - 2007. - Vol. 24. - P.621–631.

[51] Dewoody J., Nason JD, Hipkins VD Mitigating scoring errors in microsatellite data from wild populations // Molecular Ecology Notes. - 2006. - Vol. 6. - P.951–957.

[52] Dakin EE, Avise JC Microsatellite null alleles in parentage analysis // Heredity. –2004. - Vol. 93. - P.504-509.

[53] Rungi D., Berube Y., Zhang J., Ralph S., Ritland CE, Ellis B E., Douglas C., Bohlmann J., Ritland K. Robust simple sequence repeat markers for spruce (Picea spp.) From expressed sequence tags // Theor Appl Genet. –2004. - Vol. 109. - P.1283-1294.

[54] Wake DB, Wake MH, Specht CD Homoplasy: From Detecting Pattern to Determining Process and Mechanism of Evolution // Science. - 2011. - Vol. 331. - P.1032-1035.

[55] Estoup A., Jarne P., Cornuet J.-M. Homoplasy and mutation model at microsatellite loci and 55.their consequences for population genetics analysis // Molecular Ecology. - 2002. - Vol. 11. - P. 1591-1604.

[56] O'hanlon PC, Peakall R. A Simple method for the detection of size homoplasy among amplified fragment length polymorphism fragments // Molecular Ecology. - 2000. - Vol.9. - P.815-816.

[57] Vekemans X., Beauwens T., Lemaire M., Roldcn-Ruiz I. Data from amplified fragment length polymorphism (AFLP) markers show indication of size homoplasy and of a relationship between degree of homoplasy and fragment size // Molecular Ecology. - 2002. - Vol. 11. - P.139-151.

[58] Caballero A., Quesada H. Homoplasy and Distribution of AFLP Fragments: An Analysis In Silico of the Genome of Different Species // Mol. Biol. Evol. - 2010. - Vol. 27. - P.1139-1151.

[59] Herrmann D., Poncet BN, Manel S., Rioux D., Gielly. L, Taberlet P. Gugerli F. Selection criteria for scoring amplified fragment length polymorphisms (AFLPs) positively affect the reliability of population genetic parameter estimates // Genome. - 2010. - Vol. 53. - P.302-310.

[60] Whitlock R, Hipperson H, Mannarelli M.,. Butlin R .K, Burke T. An objective, rapid and reproducible method for scoring AFLP peak-height data that minimizes genotyping error // Molecular Ecology Resources. - 2008. - Vol. 8. - P.725-735.

[61] Zhang H., Hare MP Identifying and reducing AFLP genotyping error: an example of tradeoffs when comparing population structure in broadcast spawning versus brooding oysters // Heredity. –2012. - Vol. 108. - P. 616-625.