REVIEW

Trends in plant research using molecular markers

Jose Antonio Garrido-Cardenas[1] · Concepción Mesa-Valle[1] · Francisco Manzano-Agugliaro[2]

Abstract

Main conclusion **A deep bibliometric analysis has been carried out, obtaining valuable parameters that facilitate the understanding around the research in plant using molecular markers.**

The evolution of the improvement in the field of agronomy is fundamental for its adaptation to the new exigencies that the current world context raises. In addition, within these improvements, this article focuses on those related to the biotechnology sector. More specifically, the use of DNA markers that allow the researcher to know the set of genes associated with a particular quantitative trait or QTL. The use of molecular markers is widely extended, including: restriction fragment length polymorphism, random-amplified polymorphic DNA, amplified fragment length polymorphism, microsatellites, and single-nucleotide polymorphisms. In addition to classical methodology, new approaches based on the next generation sequencing are proving to be fundamental. In this article, a historical review of the molecular markers traditionally used in plants, since its birth and how the new molecular tools facilitate the work of plant breeders is carried out. The evolution of the most studied cultures from the point of view of molecular markers is also reviewed and other parameters whose prior knowledge can facilitate the approach of researchers to this field of research are analyzed. The bibliometric analysis of molecular markers in plants shows that top five countries in this research are: US, China, India, France, and Germany, and from 2013, this research is led by China. On the other hand, the basic research using Arabidopsis is deeper in France and Germany, while other countries focused its efforts in their main crops as the US for wheat or maize, while China and India for wheat and rice.

Keywords AFLP · QTL · RAPD · RFLP · SNP · Microsatellite

Abbreviations

AFLP	Amplified fragment length polymorphism
NGS	Next generation sequencing
PCR	Polymerase chain reaction
QTL	Quantitative trait loci
RAPD	Random amplification of polymorphic DNA
RFLP	Restriction fragment length polymorphism
SNP	Single-nucleotide polymorphisms
SSR	Simple sequence repeats
STR	Short tandem repeats

✉ Jose Antonio Garrido-Cardenas
   jcardena@ual.es

   Concepción Mesa-Valle
   cmesa@ual.es

   Francisco Manzano-Agugliaro
   fmanzano@ual.es

1   Department of Biology and Geology, University of Almeria, 04120 Almeria, Spain

2   Department of Engineering, University of Almeria, 04120 Almeria, Spain

Introduction

The improvement in the yield of agricultural crops in the last century has seen remarkable progress (Bohra et al. 2014). However, there are still areas for improvement. Agronomy has evolved at the same pace as social, migratory, and cultural changes have been taking place in the world; therefore, the need for new genotypes is enormous today. Plant breeders are at the crossroads of continually improving the varieties they work with to adapt to market needs, consumer demands, and growing agronomic problems (climate, pests, soil conditions, etc) (Evans 1997; Reynolds and Rodomiro 2010).

While most of the progress achieved so far has been achieved with techniques of classical improvement, future prospects go through the control of biotechnology as a

fundamental condition to obtain a greater probability of success in crop improvement (Lucht 2015). Within biotechnology, the study and use of DNA markers for plant breeding provide an encouraging picture (Lateef 2015). It should not be forgotten that many of the breeds we have pointed out that concern the agricultural sector, such as pest resistance or yield, are genetically determined. What happens is that there is usually not a single gene that uniquely determines these characters, as Mendel studied in the 19th century. Normally, there is a set of genes that are, as a whole, controlling a certain trait. The regions of the genome in which the genes associated with a particular quantitative trait are located are called QTLs, quantitative trait loci (Collard et al. 2005). That is why it is fundamental to build linkage maps and carry out QTL analysis that shows the relationship between a genomic region and its associated trait (Wang et al. 2016). This process is called QTL mapping (Broman et al. 2003).

The use of DNA markers associated with important agronomic factors is widespread in the improvement of various types of crops such as rice (*Oryza sativa*) (Mackill et al. 1999), maize (*Zea mays*) (Ortiz 2010; Suwarno et al. 2015), wheat (*Triticum aestivum*) (Landjeva et al. 2007) or tomato (*Lycopersicon esculentum*) (Illa-Berenguer et al. 2015). However, these are also being used globally to optimize efficiency in the production of other types of food, such as vegetables (Xiong et al. 2015) and pastures (Eathington et al. 2007). To this end, new approaches due to the increasing availability of data provided by the sequencing of complete genomes and transcriptomes are fundamental results. In fact, the complete genome of many species with agronomic interest such as rice (Sasaki and Burr 2000) or tomato (Tomato Genome Consortium 2012) already exists. These new technologies are offering a large amount of genomic sequences at a very low price and in a very short time (Garrido-Cardenas et al. 2017). Thus, genetic improvement is expected to benefit from this new circumstance and optimize both the efficiency and accuracy of the whole process.

The objective of this manuscript is to carry out a bibliometric study on the use of molecular markers in plants in the last 50 years. Previously, a definition and an analysis of the main types of markers traditionally used are made. New tools used to improve the identification of markers such as microarrays or massive sequencing or next generation sequencing (NGS) are also presented, and future perspectives are advanced.

Molecular markers' overview

Molecular markers have been used in recent years in the agronomic sector as powerful tools for the analysis of genetic variation as they offer an efficient way of linking phenotypic and genotypic variation (Varshney et al. 2005;

Grover and Sharma 2016). However, not all markers are equally valid. The characteristics that a good marker has to fulfil will depend, to a large extent, on the size and composition of a plant population and the number of genes segregating in a population (Collard and Mackill 2008). However, in any case, all molecular markers analysis techniques must meet the following criteria: (1) reliability. Molecular markers should be very close to an investigated locus. The results are improved using several markers if they are flanking at a loci or intragenic; (2) being highly polymorphic, to discriminate between different genotypes, and to be evenly distributed in the genome; (3) having to be a simple, cheap, and fast technique; and (4) needing very little genetic starting material to carry out the analysis.

Based on the method of analysis, molecular marker techniques can be classified into three categories: (1) non-PCR-based techniques (Lander and Botstein 1989), but based on hybridization, i.e., restriction fragment length polymorphisms (RFLPs); (2) PCR-based techniques (O'Hanlon et al. 2000). This category belongs a large number of techniques such as random amplification of polymorphic DNA (RAPD) and amplified fragment length polymorphisms (AFLP). This category, in turn, could be divided into two, depending on whether primers designed from known sequences or degenerate primers are used; and (3) sequence-based marker techniques (Ganal et al. 2009), that is, single-nucleotide polymorphisms (SNPs).

Molecular markers types

As noted above, the use of one or the other technique will depend on both the study population and the phenotype and genotype analyzed in the study. In addition, often, in a research project, the researcher is not limited to carry out a single analysis of molecular markers, but instead performs the combination of several of them (Kumar 1999). To this must be added that new techniques of DNA analysis offer a large amount of data, whose study is not yet fully normalized. Therefore, it is difficult to make a list of the different individual markers available, so that in this article will list and describe the traditionally most used.

Restriction fragment length polymorphism (RFLP)

Detection of the marker is performed by hybridization techniques, labelling a DNA fragment to be used as a probe and carrying out a Southern blot analysis (Williams 1989). What is done is to digest different DNA samples with restriction enzymes in the hope that the sequence differences will occur at the cleavage sites of these restriction enzymes, so that a different and characteristic digestion pattern is obtained of each DNA sequence. RFLP markers are usually designed to

detect both alleles in a heterozygous sample. Using this technique, they can be identified from point mutations, such as single-nucleotide polymorphisms, to DNA insertions, deletions, or rearrangements. Given the characteristics of analysis and its simplicity, through RFLP can be studied a large number of samples at a time, as well as a large number of markers in a single sample. The main drawbacks of this technique are: the application of RFLP is very time consuming; it needs relatively large and high-quality amounts of DNA of known sequence; the labelling of the probes is usually on the radioisotope P32; and the high cost of the technique.

Random-amplified polymorphic DNA (RAPD)

The objective through the use of RAPD markers is the obtaining of fragments of different sizes after carrying out a reaction of PCR on the genomic DNA that is being studied (Williams et al. 1990). In practice, what is done is to design random primers that are to be attached to different regions of the DNA, so that a given profile is to be obtained for each pair of primers. If, as a consequence of a mutation, the site to which the primer has to be attached changes, the amplification products will also change, obtaining a substantially different profile. As it is easy to understand, it is not necessary to know in advance the sequence of the DNA to which the primer is to be attached. This is the main advantage of these molecular markers against RFLP, with the main disadvantages inherent to those of the PCR reaction itself: a good quality DNA template is required, the reaction conditions must be very well established, etc. Another drawbacks of RAPDs are that most of these markers are dominant; therefore, it is not usually possible to know whether the alteration has occurred in a copy or both of the DNA (Bardakci 2001).

Amplified fragment length polymorphism (AFLP)

In a way, this type of markers can be considered as a mixture of the two previous ones. As in the RAPDs markers, a PCR amplification reaction takes place (Vuylsteke et al. 2007). The difference is that, in this case, the template is DNA that has previously been digested with restriction enzymes. The second major difference is that in AFLPs, the amplification is selective rather than random (Vos et al. 1995). As in the case of RAPDs, in this case, it is not necessary to know the sequence of the DNA to be amplified beforehand and by means of this technique a series of bands of 50–300 bp, known as fingerprints (Mueller and Wolfenbarger 1999). One of the great advantages of AFLP markers is that they are easily multiplexable, which allows them to increase their performance considerably. Their main drawback is that when a fragment with low sequence homology is presented between samples, the number of common AFLPs will be

very low and the technique is no longer useful (Janssen et al. 1996).

Microsatellites

Microsatellites, also known as short tandem repeats (STRs) or as simple sequence repeats (SSRs), are repeats of up to 100 times of simple sequences of 1–8 base pairs (Hamada and Kakunaga 1982). These elements are present in both coding and non-coding regions of all eukaryotic and prokaryotic genomes studied to date, even being present in chloroplast and mitochondrial DNA (Provan et al. 2001; Chung and Staub 2003).

The primers used in the PCR reactions for the analysis of microsatellites may be labelled with a fluorophore, with a radioactive element or lacking labelling. Depending on whether one option or another is used, the detection systems will be different and can be used from laser detection systems with automatic reading to simple agarose gels. The main advantages of this type of markers are the large number of them that exist (Adal et al. 2015) and their co-dominant inheritance, which provides—in contrast to dominant markers—the complete genetic information. That is why they are probably the most widely used molecular markers in the world labs.

Single-nucleotide polymorphisms (SNP)

A single-nucleotide polymorphism is said to exist when a single-nucleotide change (A, T, C, or G) is observed by comparing the DNA of different members of a species. These changes in a single position are used as an effective genetic marker in practically all the studied species both animal (Kim et al. 2010) and vegetal (Ganal et al. 2012), due to its great abundance, and its importance has become remarkable in the genetic analyzes in the last years. Due to their characteristics, they are extremely useful in a multitude of analyzes, being able to evaluate a large number of loci and discriminating efficiently between homozygous and heterozygous alleles. In addition, SNPs are homogeneously distributed throughout the genome, they have low mutation rates, and they show high heritability, making them ideal markers. Depending on the type of mutation that occurs, the SNPs can be classified into: (i) transversions, with changes in nucleotides C/G, A/T, C/A, and T/G; (ii) transitions, appearing C/T or G/A changes; and (iii) indels, produced by insertions or deletions of a single nucleotide. In plants, thanks to the recent development of different molecular techniques such as massive sequencing (Davey et al. 2011), it has been possible to design high-performance routine SNP analysis that allows the study of thousands of positions at a time.

New tools used in the detection of molecular markers

At present, the global needs of a world, whose population does not stop growing, demand to put new tools in the hands of the breeders (Tester and Langridge 2010). The Food and Agriculture Organization of the United Nations, FAO, already speaks of a greener revolution. Its goal is to end global malnutrition through crop science. In addition, for this, it is fundamental to use both conventional breeding techniques and the new tools that have emerged in the area of molecular genetics (Pérez-de-Castro et al. 2012). Within these tools, there are two that stand out over the others for the low price of their analyzes and for the high performance achieved in obtaining data. These are microarrays and the next generation sequencing, NGS.

Microarrays

Since the end of the 20th century, microarrays have been used, above all, to know the transcriptional activity of a biological sample (Slonim and Yanai 2009). Although other techniques were previously used in gene expression studies such as Northern blot or later quantitative PCR, the introduction of microarrays facilitated the analysis of thousands of genes at the same time in a same reaction, increasing the sensitivity and lowering the detection threshold of the transcriptional level of the less represented genes of a mixture (Kerr et al. 2000). Microarray assays are developed on a solid surface to which thousands of genomic sequences called probes have been covalently bound to be hybridized with a biological sample that has been fluorescently labelled (Heller 2002). Thereafter, each fluorescence signal will be individually detected, so that the data set obtained will result in a hybridization map. By attaching tens of thousands of DNA fragments to each support, the main advantage of using this technology lies in the high number of analyzes that can be performed in parallel. Microarrays are currently being used for a large number of assays related to gene expression, such as in the detection of a tumor profile (Pacheco-Marín et al. 2016), in the study of gene regulation in a developmental process or in the detection of mutations for the genotyping of a sample (Gunderson et al. 2005).

Next generation sequencing, NGS

Next generation sequencing is a set of techniques, whose fundamental goal is the parallelization of DNA sequencing, so that thousands or millions of molecules of genetic material can be read simultaneously (Hall 2007). There are currently up to eight large massive sequencing platforms

(Goodwin et al. 2016). Each of these platforms develops in a different way the preparation of the sample, its analysis, and the collection of the data. In any case, regardless of the technique used, massive sequencing allows the development of high-density genetic maps by identifying a large number of markers (Rasheed et al. 2017). This technology has been used successfully in the detection of SNPs of different species well known genetically like pine or maize (Eckert et al. 2009; Yan et al. 2010). Through massive sequencing, genetic maps of species not so well known as eucalyptus have also been built (Neves et al. 2011). The NGS methodology used in the field of agronomy has facilitated the identification of molecular markers linked to both QTLs and individual genes, thus optimizing the results obtained using the classical methodology (Mateo-Bonmatí et al. 2014).

Methodology

The bibliometric analysis allows the analysis of the scientific literature with the objective of throwing data on the scientific production, in a certain subject (Singh et al. 2014; Garrido-Cardenas and Manzano-Agugliaro 2017), to understand the evolution of science. Bibliometrics is presented as a very useful tool to understand the relative importance of articles published in a scientific area (Fábregas-Ruesgas et al. 2015). This study was performed after the authors conducted a complete search of the Elsevier database, Scopus, using the following query: (TITLE-ABS-KEY (molecular markers)) AND (TITLE-ABS-KEY (plants)). The search range focused on the period 1967–2016. It should be noted that if any of the parameters in the search is altered, the results obtained may be very different. The above general search query has been completed with specific search queries, e.g., in searches of the number of documents by countries referring to types of molecular markers such as, for France and RAPD, it was: (TITLE-ABS-KEY (molecular AND markers)) AND (TITLE-ABS-KEY (plants)) AND (LIMIT-TO (AFFILCOUNTRY, "France ")) AND (LIMIT-TO (EXACT-KEYWORD, "Random Amplified Polymorphic DNA")). Another example for each crop, the common and scientific name was taken into account, and also the botanical genus, for example, for US and wheat, it was: (TITLE-ABS-KEY (molecular markers)) AND (TITLE-ABS-KEY (plants)) AND (LIMIT-TO (AFFILCOUNTRY, "United States ")) AND (LIMIT-TO (EXACTKEYWORD, "Triticum Aestivum") OR LIMIT-TO (EXACTKEYWORD, "Wheat") OR LIMIT-TO (EXACTKEYWORD, "Triticum")). This procedure ensures that one publication is counted only once.

The overlap of main scientific databases and their impact of using different data sources for specific research fields on bibliometric indicators have been measured in some studies. Therefore, they conclude that Scopus citations are

comparable to *Web of Science* citations when limiting the citation period to 1996 and onwards. Both databases cover about 90% of the citations of the other, respectively (Gimenez and Manzano-Agugliaro 2017; Salmerón-Manzano and Manzano-Agugliaro 2017). In the regard of the journal coverage, a *Web of Science* and *Scopus* comparative analysis shows that the coverage of active scholarly journals in WoS (13,605 journals) is lower than Scopus (20,346 journals) (Mongeon and Paul-Hus 2016), and the correlations between the measures obtained with both databases for the number of papers and the number of citations received by countries, as well as for their ranks, are extremely high ($R^2 \approx 0.99$) (Archambault et al. 2009). The advantages of Scopus for bibliometric analysis are shown in several research papers (Montoya et al. 2014, 2017).

The data obtained after the query of the database were processed using spreadsheets. To facilitate the visualization of the results and optimize the development of the analysis, the corresponding graphs were generated from the data obtained. The aspects studied were: (1) number of publications per year; (2) categories of distribution issues and journals; (3) type of document and language; (4) distribution by country and institution; and (5) keywords.

Results

Evolution of scientific output

The search yielded 20,794 results, whose evolution is represented in Fig. 1. It can be observed that until the end of the 80s of the 20th century, there is no remarkable growth, registering only 98 documents in the first 20 years. However, from this moment, the growth is constant until today, adjusting a second-order polynomial growth with a correlation coefficient of $R^2 = 0.9505$. The maximum number of published annual papers on molecular markers was 1744 and it was reached in 2014.

To deepen the analysis of the evolution of scientific production in this field, Fig. 2 has been made. It shows the publications trends for the top five countries. It can be observed how Top 1, the US, leaves to lead this worldwide scientific research on 2013 when it begins to be led by China. On the other hand, a constant trend is maintained by Germany and France, but India, since 2008 seems to take off in this research field, going on to maintain the third place since then; therefore, in the last year of study, it is at the same level as the US on number of publications.

Distribution of output in subject categories and journals

In the analysis of the distribution of publications by field, it should be noted that each article can be indexed in more than one category. Figure 3 shows the areas with more than 100 publications in the period studied. The analysis was carried out according to the classification of *Scopus*, and it can be observed that the first two places of this classification correspond to the categories of *Agricultural and Biological Sciences* and *Biochemistry, Genetics and Molecular Biology*, with 13,041 and 12,956 publications, respectively. These two areas together represent around 90% of all publications.

Fig. 1 Publication trends from 1985 to 2016 on molecular markers in plants
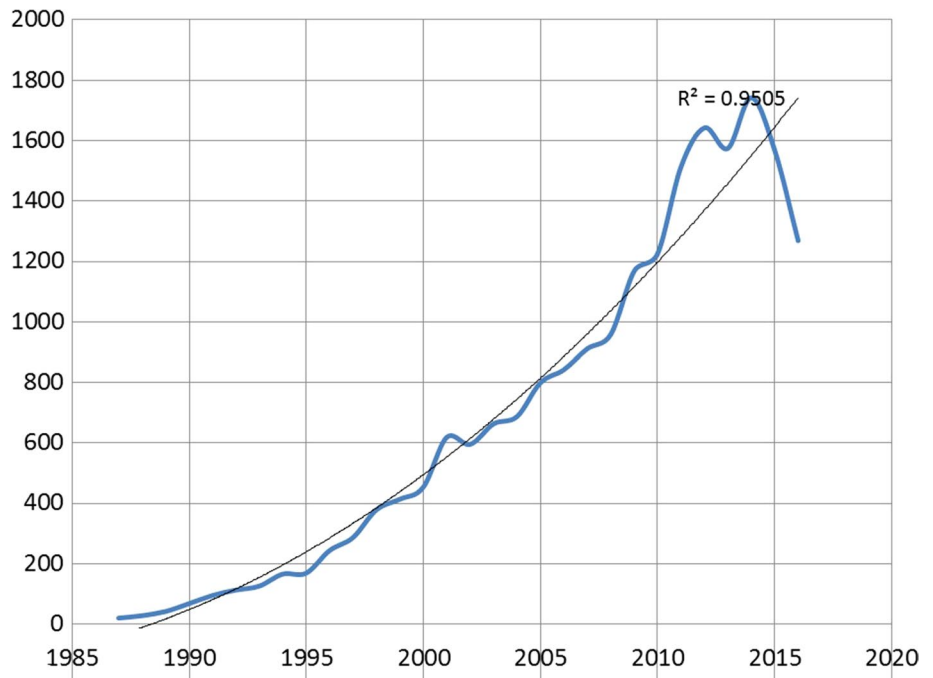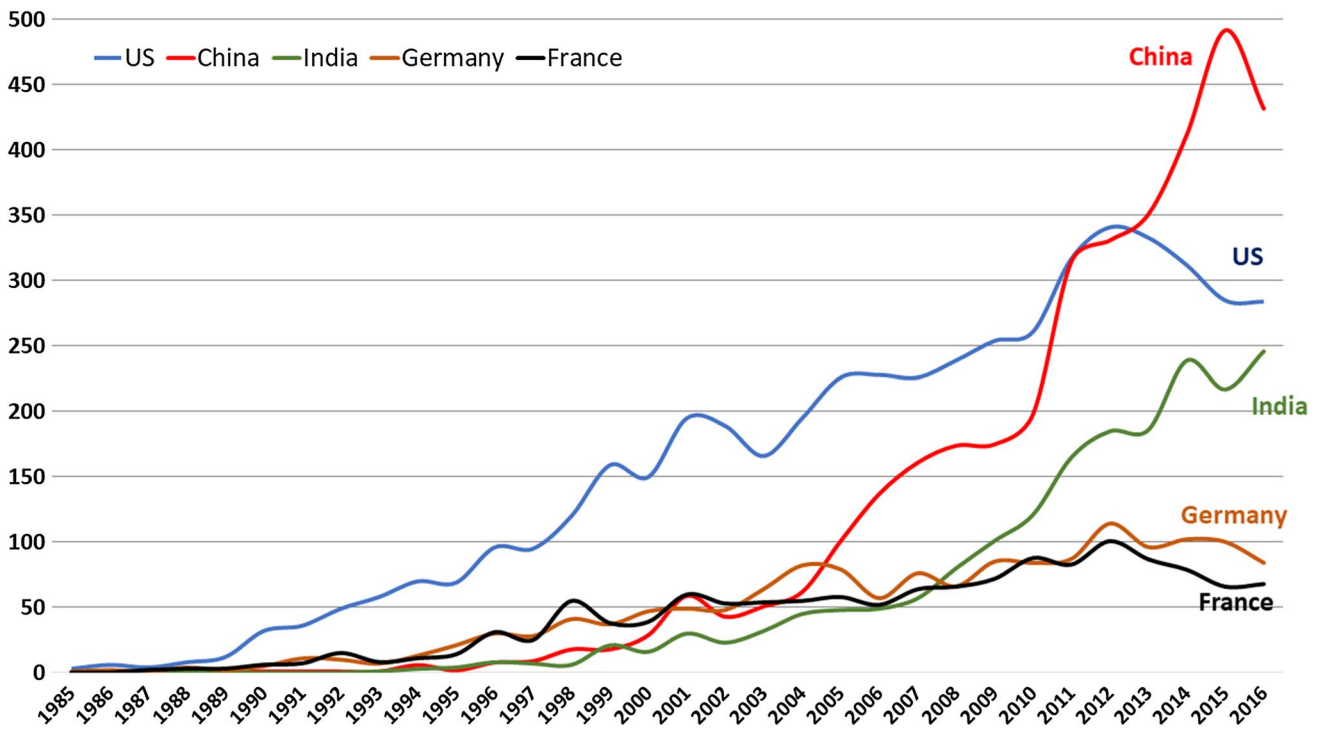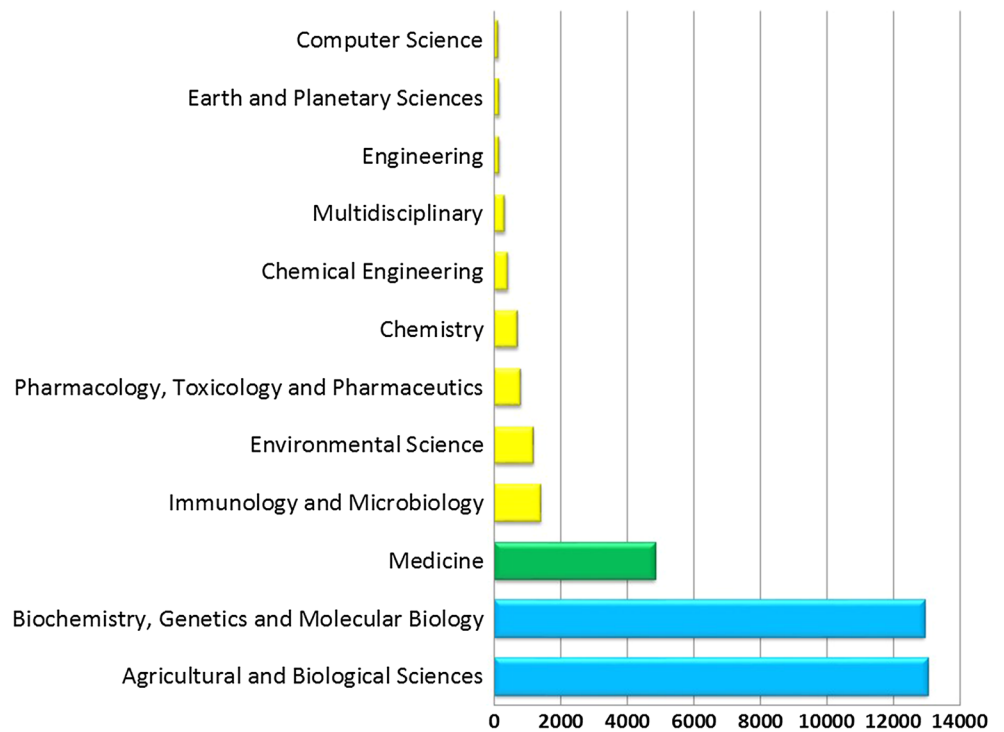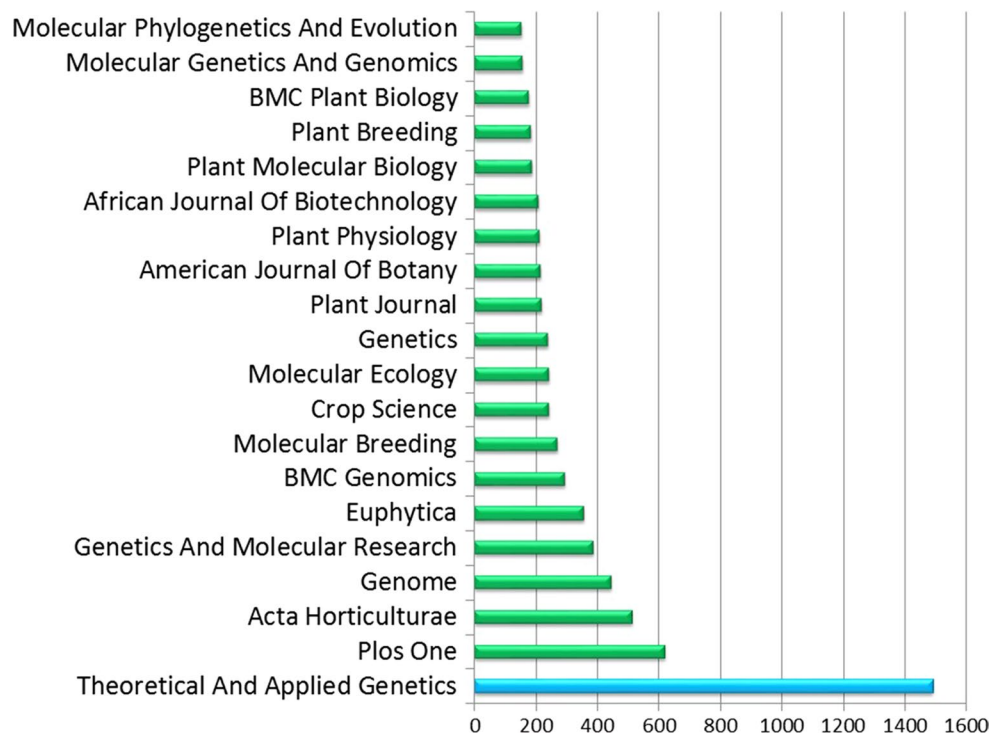
Fig. 2 Publications trends for the top five countries

Fig. 3 Distribution of publications by field

Fig. 4 Distribution of publications by source

Then, at a considerable distance, the area of *Medicine* appears, and in the fourth and fifth positions are the areas *Immunology and Microbiology* and *Environmental Science*, with just over 1000 publications each.

Figure 4 shows the journals with the highest number of publications on molecular markers in plant in the period 1967–2016. The graph only shows the journals that in this period have published at least 150 articles, resulting in a total of 20 journals. Of these, six journals are from US, four journals are from UK, three journals are from Germany, and three journals from The Netherlands. Four other countries publish a single journal: Belgium, Canada, Brazil, and Kenya. Leading this ranking stands out the journal *Theoretical and Applied Genetics*, with 1492 articles (more than the sum of the two journals that occupy the second and third positions, *Plos One* and *Acta Horticulturae*). *Theoretical and Applied Genetics* has been the journal that more articles have published in all the historical series until year 2011. From that moment, the journals *Plos One* and *Genetics and Molecular Research* have moved to lead the classification.

Types and language of publications

Figure 5 shows the type of documents that have been published in the studied period. As can be seen, the clear majority of these are articles. 18,310 publications represent 88.05% of the total. With much less representation are the reviews (1190 publications, 5.72%), the conference paper (671 publications, 3.23%), and the book chapter (358 publications, 1.72%). The rest of the publications have a mere testimonial representation and do not reach 1% of the total of the publications.

On the other hand, since English is the prevailing language in international journals, it is not surprising that 96.29% of the articles are written in this language. Languages such as Chinese (2.14%), Portuguese (0.59%), Russian (0.59%), or Spanish (0.36%) appear behind them, but with a very minority presence.

Distribution of publications by country and institutions

Figure 6 represents a world map in which countries with at least 1000 publications on molecular markers in plants are colored in brown and red. Above all, USA and China stand out, being the only two countries with more than 2000 items, specifically, 4975 articles in USA and 3470 in China, during the studied period. These two countries together publish, practically, 40% of all the articles of this subject. The remaining countries with at least 1000 publications are India (1847 articles), Germany (1532 articles), France (1342 articles), UK (1239 articles), Italy (1060 articles), and Japan (1058 articles). In the same sense, the three institutions that publish the most articles according to the search are of American or Chinese nationality. These institutions are: the USDA Agricultural Research Service, Washington DC, with 483 publications,
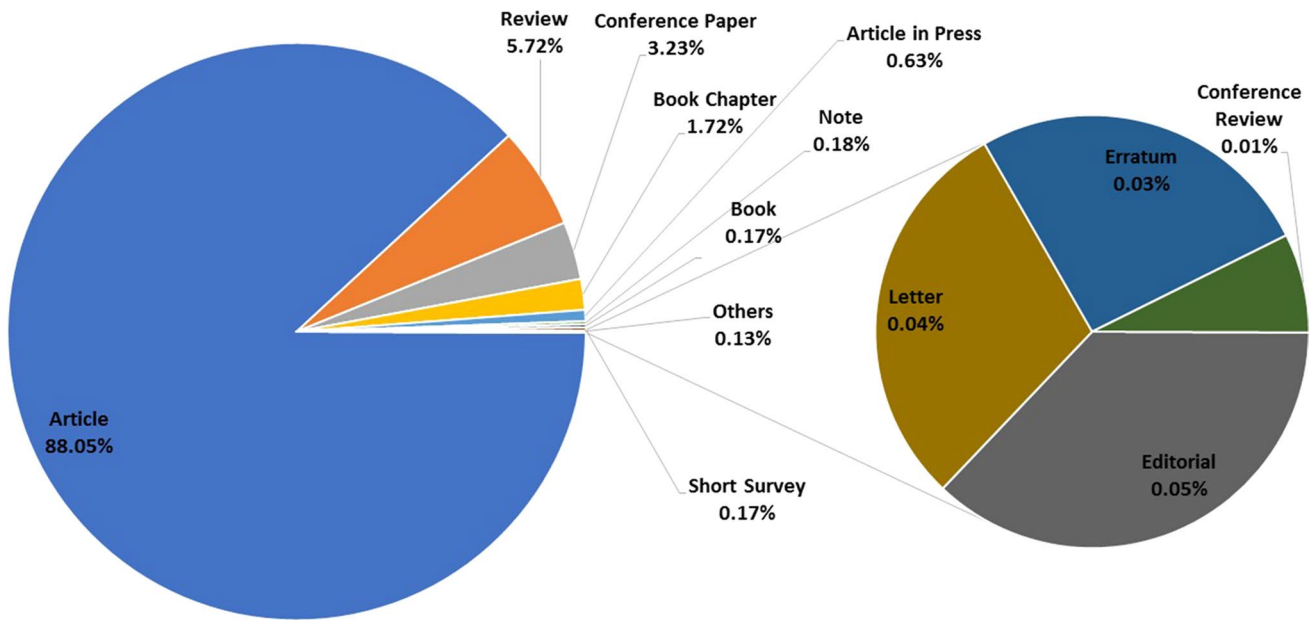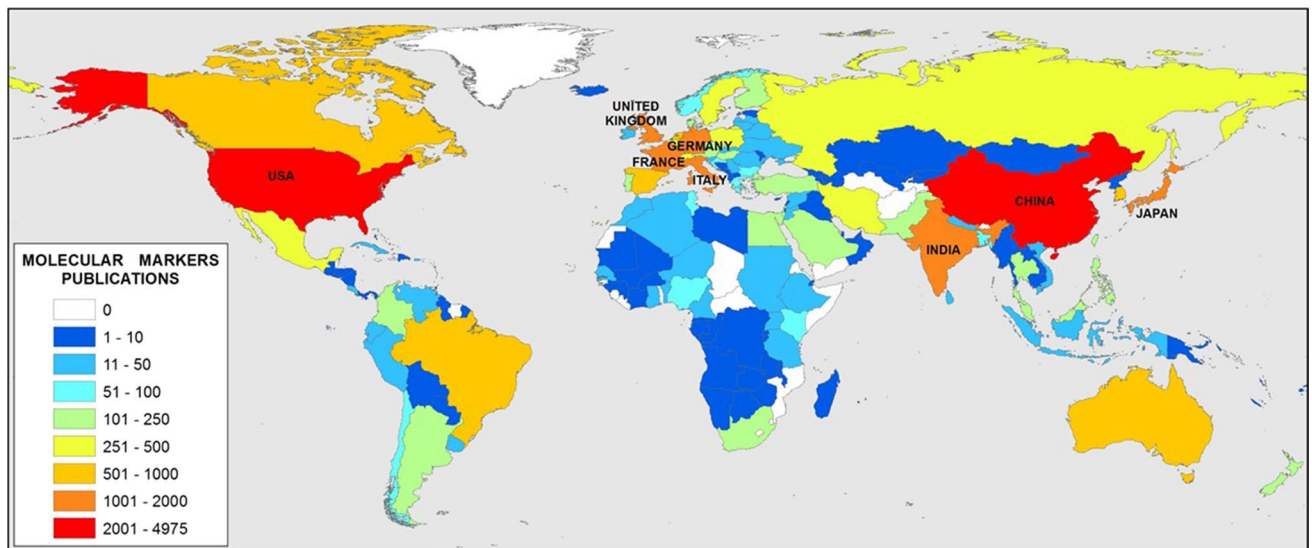
Fig. 5 Distribution of document types

**Fig. 6** World map representing the molecular markers in plants publications

the Chinese Academy of Agricultural Sciences, with 432 publications, and the University of California, UC Davis, with 338 articles. Figure 7 shows the 11 institutions with at least 200 publications on molecular markers in plants. Of these 11 institutions, four are American and four are Chinese. The other nationalities represented are: Dutch, with Wageningen University and Research Center; German, with Leibniz Institute of Plant Genetics and Crop Plant Research; and Brazilian, with the Brazilian Agricultural Research Company—Embrapa.

Keyword analysis

To carry out the analysis of the keywords, two additional adjustments had to be made in the search. On one hand, generic terms, like "article", which do not contribute anything to the study, were eliminated. In addition, on the other hand, the terms that referred to the same concept, but appeared independently, as in the case of "Plant DNA" and "DNA, Plant" or "Nucleotide sequence" and "base sequence", were grouped. Only then does it make sense to analyze the keywords to try to understand the research trends

Fig. 7 Main institutions in molecular markers plants publications
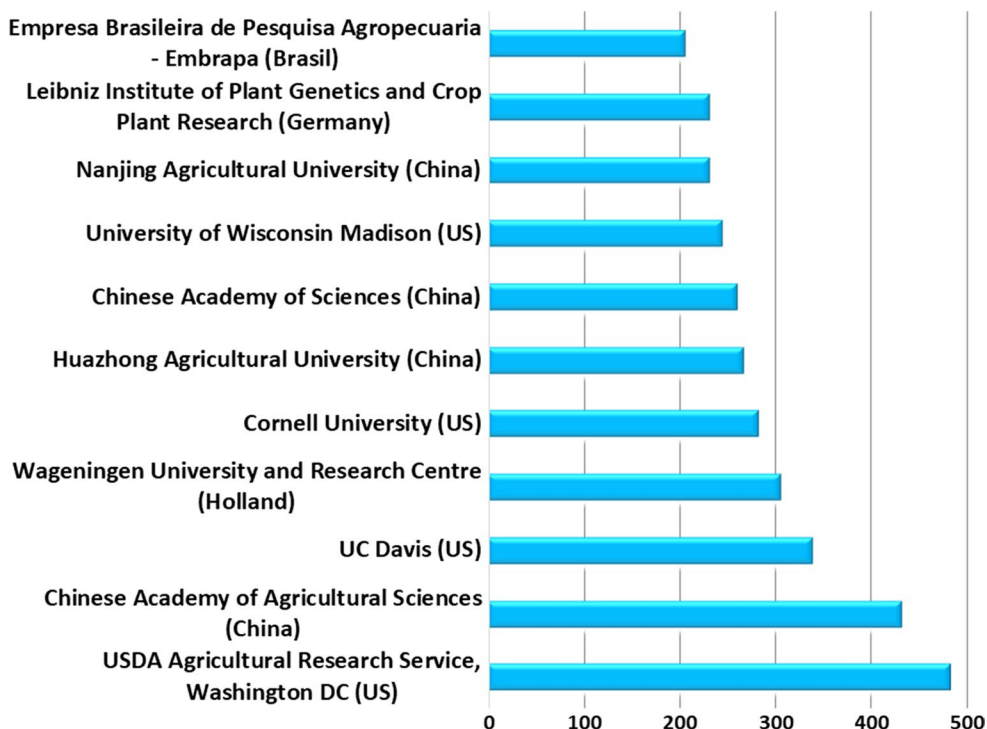
Fig. 8 Distribution of keywords
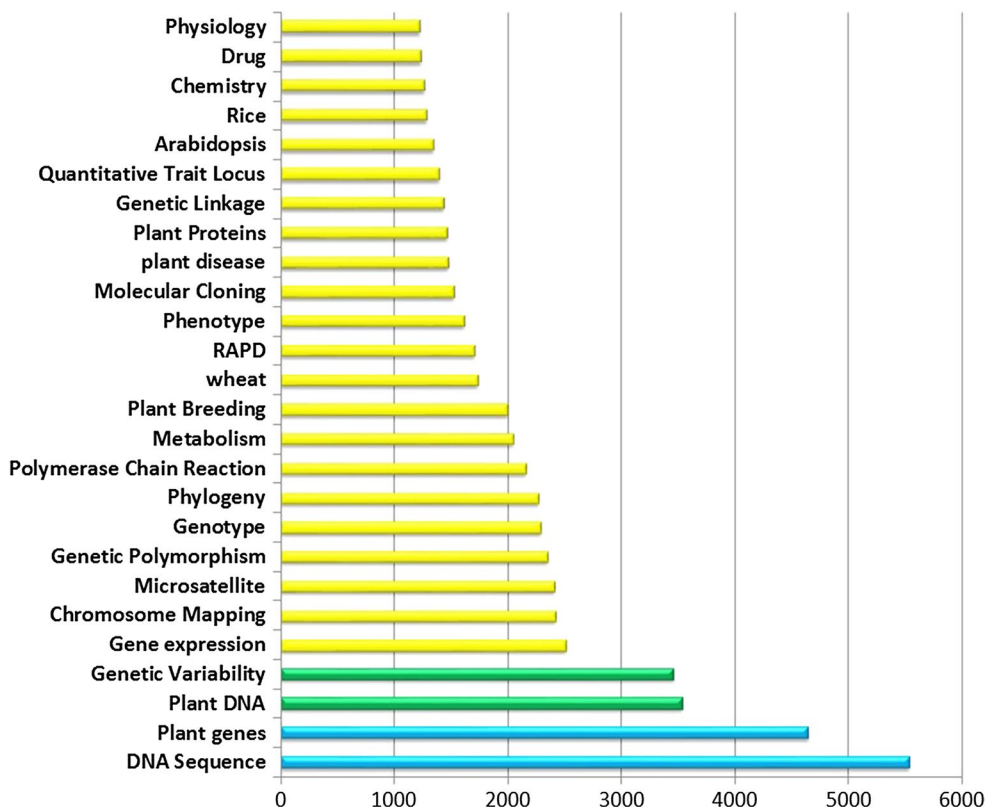
that are developed in a given area (Choi 2011). After making these two adjustments, it has been seen that there are 26 terms that appear in at least 1200 publications (Fig. 8). Note that the number of keywords that appears in each publication is variable, as it depends on each journal, and usually varies between 4 and 8.
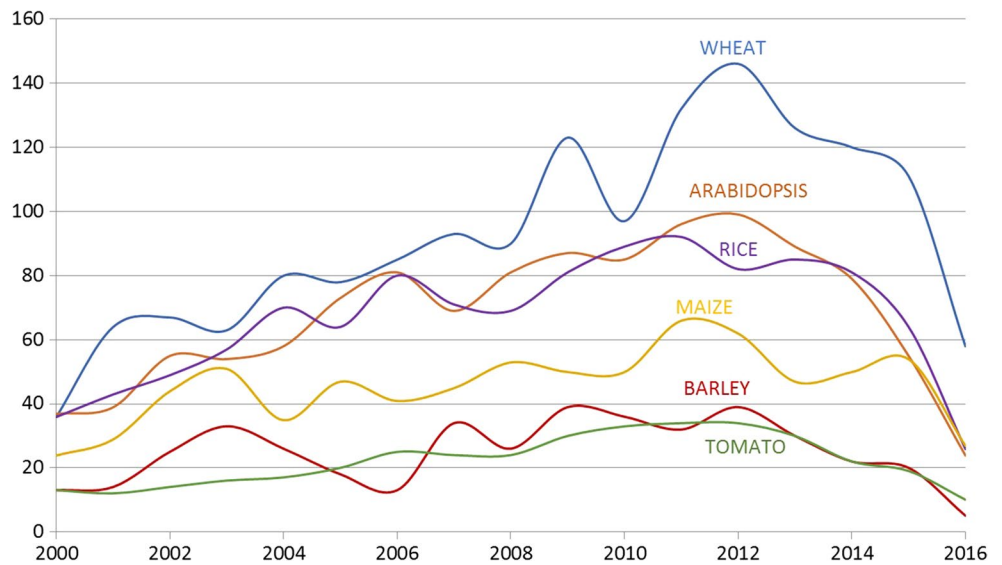
Fig. 9 Word cloud of worldwide research in molecular markers in plants

In addition, the representation of these 26 keywords in a cloud word (Fig. 9), where the number of times a keyword appears in publications with their size in the cloud, is represented proportionally. This image gets to offer a more visual analysis result.

On the other hand, an analysis of the evolution of two different series of keywords between the years 2000 and 2016 has been carried out. These two series are: (i) types of molecular markers (Fig. 10) and (ii) cultivable plant species (Fig. 11). In the first of the series, the keywords that appear are: microsatellite, RAPD, AFLP, SNP, and RFLP. In addition, in the second of the series, the keywords that appear

are: wheat, Arabidopsis (*Arabidopsis thaliana*), rice, maize, barley (*Hordeum vulgare*), and tomato.

In Fig. 10 the appearance of RAPD and AFLP techniques as keywords is relatively constant in this period, whereas it is not in the rest of the techniques, with a significant decrease in the RFLP technique and a considerable increase in the presence of microsatellites and SNPs as keywords. In absolute terms, the methodology with a greater presence among the keywords in the studied years is microsatellite, followed by RAPD, whereas the one that counts with a smaller presence is the RFLP technique, reinforcing the specific weight loss that it is suffering in the last years.

Fig. 10 Evolution of main keywords related to molecular markers from 2000 to 2016

Fig. 11 Evolution in cultivable plant species as keywords between 2000 and 2016

About the plant species that appear among the keywords (Fig. 11), in absolute terms, the one that appears with a higher frequency is wheat. The second position appears Arabidopsis, the model organism of choice for research in plant biology (Koornneef and Meinke 2010). The third and fourth places are occupied by two species of cereal, maize, and rice. These are two of the most consuming species in the world and contribute most to human consumption. These are also of great importance in animal feed, especially maize. Barley appears in the fifth place, probably because it is not only used directly in human and animal feed, but also, because it is the main component of beer, widely spread throughout the planet. Finally, in the sixth, one appears the tomato, that is one of the horticultural plants with more diffusion worldwide, as much for its volume of consumption in fresh as for its commercial commercialization in sauce.

Finally, this study of keywords must be completed with those most used by the main institutions and countries dedicated to this field. Table 1 lists the three main keywords used by the institutions as well as the keyword of the most cultivated plant. Overall, there is a lot of similarity and repetition with the keywords Nonhuman and Chromosome Mapping, which usually occupy the first two positions in almost all the research institutions. The specialization is found more when the main plant keyword that appears is selected. The first three institutions in the ranking are centered in wheat, while in the others, each one is specialized in a crop, generally related to its agricultural environment. For example, the University of Wisconsin Madison studies mainly *Cucumis Sativus*, and its species of pickle (Pickling Cucumber Wisconsin SMR 18) is well known, or the Nanjing Agricultural University that has many studies of the cotton thanks to its Cotton Research Institute, or the Genetic Research and Breeding of Rapeseed at Huazhong Agricultural University dedicated to rapeseed (*Brassica napus*).

Table 1 Top three keywords and main plant keyword for the top ten institutions

| Affiliation	Main keywords			
	1	2	3	Plant
USDA Agricultural Research Service, Washington, DC	Chromosome mapping	Nonhuman	Plant diseases	Wheat
Chinese Academy of Agricultural Sciences	Nonhuman	Chromosome mapping	Genes, plant	Wheat
UC Davis	Nonhuman	Chromosome mapping	Genotype	Wheat
Wageningen University and Research Centre	Nonhuman	Chromosome mapping	Genotype	Potato/tomato
Cornell University	Nonhuman	Chromosome mapping	Molecular sequence data	Tomato/rice
Huazhong Agricultural University	Nonhuman	Chromosome mapping	*Brassica napus*	Rapeseed
Chinese Academy of Sciences	Nonhuman	Phylogeny	Molecular sequence data	Rice
University of Wisconsin Madison	Nonhuman	Chromosome mapping	Genome, plant	Cucumber
Nanjing Agricultural University	Nonhuman	Chromosome mapping	Chromosomes, plant	Cotton
Leibniz Institute of Plant Genetics and Crop Plant Research	Hordeum	Nonhuman	Barley	Barley
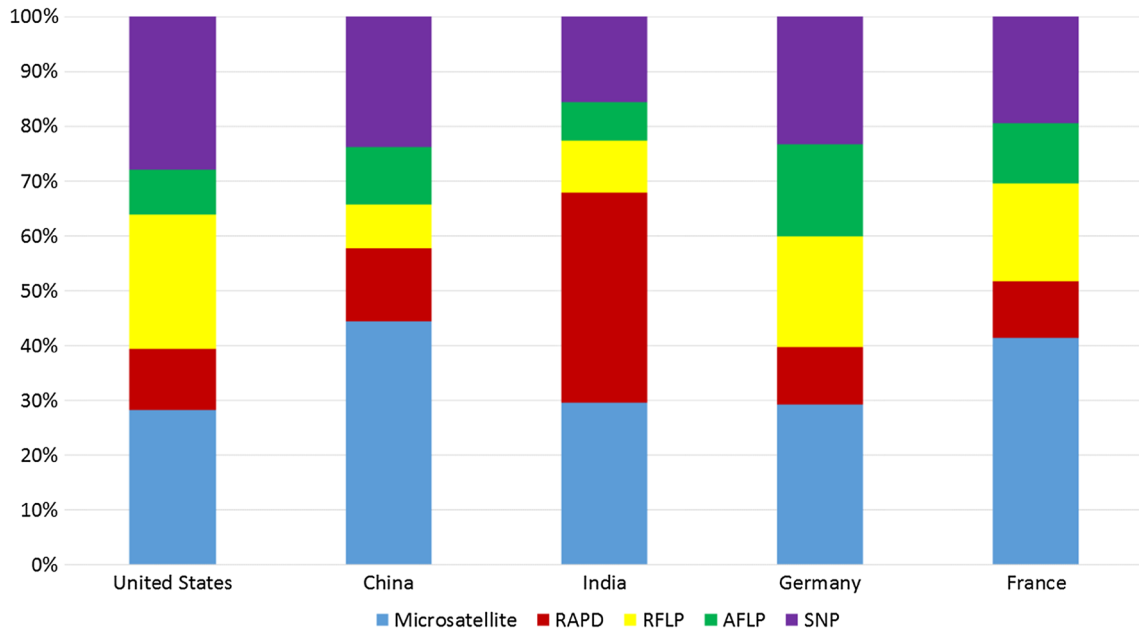
Fig. 12 Main keywords of molecular markers for top five countries
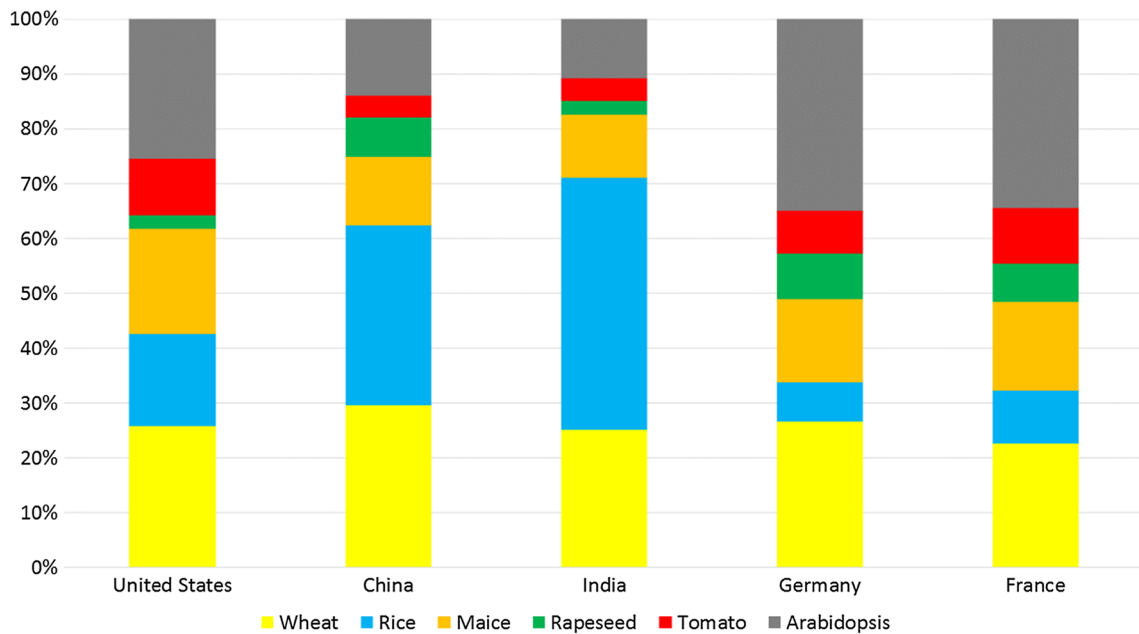

Fig. 13 Main keywords of plants for top five countries

If one pay attention to the main keywords by country, and it is distinguished by molecular markers and by the crops studied, we obtain Figs. 12 and 13, respectively. The representation attends the percentage of these publications among them. As shown in Fig. 12, the relative importance of RFLP in USA is greater than in the other studied countries, while microsatellite is for China and France or RAPD is for India. On the other hand, the keywords of the crops for each country show the main interest of each country in them. Thus, it is observed how the experimentation with plants for the basic research such as Arabidopsis rounds 35% in France and Germany, while the US is about 25%; in this aspect, China and India are below 15%. Regarding wheat, all countries have a high interest, between 22% for France

and 30% for China. The largest variations are found related to the rice, where France and Germany are below 10%, while India and China reach values of 33–46%, respectively. In the study of corn, US stands out with almost 20% of its publications dedicated to it. Regarding rapeseed, having all values below 10%, Germany stands out with 8% and China with 7%. Finally, for tomato, highlights the interest of two countries with values above 10%, US and France.

Conclusions and discussion

The publications about the theme of molecular markers in plants between the years 1967 and 2016 have been analyzed. It has started from the search carried out in the Scopus database, and the aspects that have been studied are: evolution of scientific output, distribution of output in subject categories and journals, types and languages of publications, distribution of publications by country and institutions, and keyword analysis. The first thing to note is that the number of publications grew very moderate, since the first articles appeared until the last decades of the 20th century, but since then, the number of articles published has not stopped growing and it has done it following a polynomial function of the second order. It is also noted that the clear majority—96%—are written in English, 88% are articles and 90% are classified under the categories *Agricultural and Biological Sciences* and *Biochemistry, Genetics and Molecular Biology*.

The top five countries are: US, China, India, Germany, and France. It is emphasized that USA dominates this field until 2013, since it begins to be led by China. On the other hand, India has advanced a lot in this field, arriving in the last year of study at the US level. Between them publish about 40% of all the articles of molecular markers in plants, being the institutions of these nationalities the most active. In fact, one US institution—the USDA Agricultural Research Service, Washington DC—and another Chinese— the Chinese Academy of Agricultural Sciences—are the ones that have published the most documents during the studied period. Showing these institutions, the interest of their country for specific crops, this can be observed, since there are research centers associated with these institutions focused on these crops.

When analyzing the evolution in the keywords of the presence of the various techniques to identify molecular markers, it shows the trend that has been kept worldwide in the use of these techniques. While the late 20th century techniques based on the use of restriction enzymes and the subsequent analysis of the fragments obtained in gels were the predominant ones, with the advent of massive sequencing technologies, the trend changed. The tedious montages evolved and the techniques were automated. New methodologies to obtain a maximum number of data

in an absolutely standardized way and analysis are carried out by software more and more specific and versatile every day. Thus, currently, analyzes as the RFLPs are merely testimonial, whereas there is an absolute predominance by the identification of SNPs and the analysis of microsatellites.

Related to specific plant studied, as was expected, cereals as wheat, maize, rice, and barley are the most studied plants by these techniques. An inedible plant such as Arabidopsis also occupies an important place because of its condition as a model plant. The most studied horticultural plant is the tomato, as expected, being the most consumed horticultural species in the world, both fresh and in sauce. Therefore, the basic research using Arabidopsis is deeper in France and Germany, while the other countries focused its efforts in their main crops as the US for wheat or maize, while China and India for wheat and rice. The study of tomato is especially important in US and France.

If the new global perspective of the world has changed, the way we interact with each other, therefore, has our way of understanding our diet. If we add to this a world population that continues to grow, it is easy to conclude that demands for wheat, rice, or maize crops will become increasingly demanding. That is why it is essential to have tools aimed at optimizing the different agronomic resources. In addition, this is where the role of molecular markers is fundamental. This manuscript shows how different molecular markers as RFLP, RAPD, and AFLP are practically not currently used; therefore, the trend of their use over time has been observed. Identifying, knowing, and manipulating genes that determine certain characteristics seem to be the only way to maximize the yield of agricultural crops.

Author contribution statement JAG-C designed the analysis, wrote the introduction part related to New tools used in the detection of molecular markers, and the part related to Molecular markers types, carried out the analysis related to "Keyword Analysis", and elaborated the Conclusions and discussion.

CM-V wrote part of the introduction, performed the analyzes related to "Evolution of scientific output", "Distribution of output in subject categories and journals", and "Types and language of publications", and elaborated the Conclusions and discussion.

FM-A designed the analysis, wrote and designed the methodology, carried out the analyzes related to "Distribution of publications by country and institutions", and elaborated the Conclusions and discussion.

Compliance with ethical standards

References

Adal AM, Demissie ZA, Mahmoud SS (2015) Identification, validation and cross-species transferability of novel Lavandula EST-SSRs. Planta 241:987–1004. https://doi.org/10.1007/s00425-014-2226-8

Archambault É, Campbell D, Gingras Y, Larivière V (2009) Comparing bibliometric statistics obtained from the Web of Science and Scopus. J Am Soc Inf Sci Technol 60:1320–1326. https://doi.org/10.1002/asi.21062

Bardakci F (2001) Random amplified polymorphic DNA (RAPD) markers. Turk J Biol 25:185–196

Bohra A, Pandey MK, Jha UC et al (2014) Genomics-assisted breeding in four major pulse crops of developing countries: present status and prospects. Theor Appl Genet 127:1263–1291

Broman KW, Wu H, Sen Ś, Churchill GA (2003) R/qtl: QTL mapping in experimental crosses. Bioinformatics 19:889–890. https://doi.org/10.1093/bioinformatics/btg112

Chung SM, Staub JE (2003) The development and evaluation of consensus chloroplast primer pairs that possess highly variable sequence regions in a diverse array of plant taxa. Theor Appl Genet 107:757–767. https://doi.org/10.1007/s00122-003-1311-3

Collard BCY, Mackill DJ (2008) Marker-assisted selection: an approach for precision plant breeding in the twenty-first century. Philos Trans R Soc Lond B Biol Sci 363:557–572. https://doi.org/10.1098/rstb.2007.2170

Choi J, Yi S, Lee KC (2011) Analysis of keyword networks in MIS research and implications for predicting knowledge evolution. Inf Manag 48:371–381. https://doi.org/10.1016/j.im.2011.09.004

Collard BCY, Jahufer MZZ, 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 142:169–196

Davey JW, Hohenlohe PA, Etter PD et al (2011) Genome-wide genetic marker discovery and genotyping using next-generation sequencing. Nat Rev Genet 12:499–510. https://doi.org/10.1038/nrg3012

Eathington SR, Crosbie TM, Edwards MD et al (2007) Molecular markers in a commercial breeding program. Crop Sci 47. https://doi.org/10.2135/cropsci2007.04.0015IPBS

Eckert AJ, Pande B, Ersoz ES et al (2009) High-throughput genotyping and mapping of single nucleotide polymorphisms in loblolly pine (Pinus taeda L.). Tree Genet Genomes 5:225–234. https://doi.org/10.1007/s11295-008-0183-8

Evans LT (1997) Adapting and improving crops: the endless task. Philos Trans R Soc Lond Ser B Biol Sci 352:901–906. https://doi.org/10.1098/rstb.1997.0069

Fábregas-Ruesgas JJ, Hernández-Abad F, Hernández-Abad V, Rojas-Sola JI (2015) Comparative analysis of a university learning experience: classroom mode versus distance mode. Int J Interact Des Manuf 9:145–157. https://doi.org/10.1007/s12008-014-0205-8

Ganal MW, Altmann T, Röder MS (2009) SNP identification in crop plants. Curr Opin Plant Biol 12:211–217

Ganal MW, Polley A, Graner EM et al (2012) Large SNP arrays for genotyping in crop plants. J Biosci 37:821–828. https://doi.org/10.1007/s12038-012-9225-3

Garrido-Cardenas JA, Manzano-Agugliaro F (2017) The metagenomics worldwide research. Curr Genet. https://doi.org/10.1007/s00294-017-0693-8

Garrido-Cardenas JA, Garcia-Maroto F, Alvarez-Bermejo JA, Manzano-Agugliaro F (2017) DNA sequencing sensors: an overview. Sensors (Basel) 17(3):1–15. https://doi.org/10.3390/s17030588

Gimenez E, Manzano-Agugliaro F (2017) DNA damage repair system in plants: a worldwide research update. Genes (Basel) 8:299. https://doi.org/10.3390/genes8110299

Goodwin S, McPherson JD, McCombie WR (2016) Coming of age: ten years of next-generation sequencing technologies. Nat Rev Genet 17:333–351. https://doi.org/10.1038/nrg.2016.49

Grover A, Sharma PC (2016) Development and use of molecular markers: past and present. Crit Rev Biotechnol 36:290–302. https://doi.org/10.3109/07388551.2014.959891

Gunderson KL, Steemers FJ, Lee G et al (2005) A genome-wide scalable SNP genotyping assay using microarray technology. Nat Genet 37:549–554. https://doi.org/10.1038/ng1547

Hall N (2007) Advanced sequencing technologies and their wider impact in microbiology. J Exp Biol 210:1518–1525. https://doi.org/10.1242/jeb.001370

Hamada H, Kakunaga T (1982) Potential Z-DNA forming sequences are highly dispersed in the human genome. Nature 298:396–398. https://doi.org/10.1038/298396a0

Heller MJ (2002) DNA microarray technology: devices, systems, and applications. Annu Rev Biomed Eng 4:129–153. https://doi.org/10.1146/annurev.bioeng.4.020702.153438

Illa-Berenguer E, Van Houten J, Huang Z, van der Knaap E (2015) Rapid and reliable identification of tomato fruit weight and locule number loci by QTL-seq. Theor Appl Genet. https://doi.org/10.1007/s00122-015-2509-x

Janssen P, Coopman R, Huys G et al (1996) Evaluation of the DNA fingerprinting method AFLP as a new tool in bacterial taxonomy. Microbiology 142:1881–1893. https://doi.org/10.1007/s00296-011-2085-6

Kerr MK, Martin M, Churchill GA (2000) Analysis of variance for gene expression microarray data. J Comput Biol 7:819–837. https://doi.org/10.1089/10665270050514954

Kim JJ, Han BG, Lee HI et al (2010) Development of SNP-based human identification system. Int J Legal Med 124:125–131. https://doi.org/10.1007/s00414-009-0389-9

Koornneef M, Meinke D (2010) The development of Arabidopsis as a model plant. Plant J 61:909–921. https://doi.org/10.1111/j.1365-313X.2009.04086.x

Kumar LS (1999) DNA markers in plant improvement: an overview. Biotechnol Adv 17:143–182

Lander ES, Botstein S (1989) Mapping mendelian factors underlying quantitative traits using RFLP linkage maps. Genetics 121:185. https://doi.org/10.1038/hdy.2014.4

Landjeva S, Korzun V, Börner A (2007) Molecular markers: actual and potential contributions to wheat genome characterization and breeding. Euphytica 156:271–296

Lateef DD (2015) DNA marker technologies in plants and applications for crop improvements. J Biosci Med 3:7–18. https://doi.org/10.4236/jbm.2015.35002

Lucht JM (2015) Public acceptance of plant biotechnology and GM crops. Viruses 7:4254–4281

Mackill DJ, Nguyen HT, Zhang J (1999) Use of molecular markers in plant improvement programs for rainfed lowland rice. Field Crop Res 64:177–185. https://doi.org/10.1016/S0378-4290(99)00058-1

Mateo-Bonmatí E, Casanova-Sáez R, Candela H, Micol JL (2014) Rapid identification of angulata leaf mutations using next-generation sequencing. Planta 240:1113–1122. https://doi.org/10.1007/s00425-014-2137-8

Mongeon P, Paul-Hus A (2016) The journal coverage of Web of Science and Scopus: a comparative analysis. Scientometrics 106:213–228. https://doi.org/10.1007/s11192-015-1765-5

Montoya FG, Montoya MG, Gómez J et al (2014) The research on energy in Spain: a scientometric approach. Renew Sustain Energy Rev 29:173–183

Montoya FG, Alcayde A, Baños R, Manzano-agugliaro F (2017) Telematics and informatics a fast method for identifying worldwide scientific collaborations using the Scopus database. Telemat Inf. https://doi.org/10.1016/j.tele.2017.10.010

Mueller UG, Wolfenbarger LL (1999) AFLP genotyping and finger-printing. Trends Ecol Evol 14:389–394. https://doi.org/10.1016/S0169-5347(99)01659-6

Neves LG, Mc Mamani E, Alfenas AC et al (2011) A high-density transcript linkage map with 1,845 expressed genes positioned by microarray-based single feature polymorphisms (SFP) in eucalyptus. BMC Genomics 12:189. https://doi.org/10.1186/1471-2164-12-189

O'Hanlon PC, Peakall R, Briese DT (2000) A review of new PCR-based genetic markers and their utility to weed ecology. Weed Res 40:239–254

Ortiz R (2010) Molecular plant breeding. Crop Sci 50:2196–2197. https://doi.org/10.2135/cropsci2010.12.0004br

Pacheco-Marín R, Melendez-Zajgla J, Castillo-Rojas G et al (2016) Transcriptome profile of the early stages of breast cancer tumoral spheroids. Sci Rep 6:23373. https://doi.org/10.1038/srep23373

Pérez-de-Castro AM, Vilanova S, Cañizares J et al (2012) Application of genomic tools in plant breeding. Curr Genomics 13:179–195. https://doi.org/10.2174/138920212800543084

Provan J, Powell W, Hollingsworth PM (2001) Chloroplast microsatellites: new tools for studies in plant ecology and evolution. Trends Ecol Evol 16:142–147

Rasheed A, Hao Y, Xia X et al (2017) Crop breeding chips and genotyping platforms: progress, challenges, and perspectives. Mol Plant 10:1047–1064

Reynolds MP, Rodomiro O (2010) Adapting crops to climate change: a summary. Clim Chang Crop Prod 51:1–8. https://doi.org/10.1079/9781845936334.0000

Salmerón-Manzano E, Manzano-Agugliaro F (2017) Worldwide scientific production indexed by Scopus on labour relations. Publications 5:25. https://doi.org/10.3390/publications5040025

Sasaki T, Burr B (2000) International rice genome sequencing project: the effort to completely sequence the rice genome. Curr Opin Plant Biol 3:138–141

Singh V, Perdigones A, García JL et al (2014) Analysis of worldwide research in the field of cybernetics during 1997–2011. Biol Cybern 108:757–776. https://doi.org/10.1007/s00422-014-0617-3

Slonim DK, Yanai I (2009) Getting started in gene expression microarray analysis. PLoS Comput, Biol, p 5

Suwarno WB, Pixley KV, Palacios-Rojas N et al (2015) Genome-wide association analysis reveals new targets for carotenoid biofortification in maize. Theor Appl Genet 128:851–864. https://doi.org/10.1007/s00122-015-2475-3

Tester M, Langridge P (2010) Breeding technologies to increase crop production in a changing world. Science (80) 327:818–822. https://doi.org/10.1126/science.1183700

Tomato Genome Consortium (2012) The tomato genome sequence provides insights into fleshy fruit evolution. Nature 485:635–641. https://doi.org/10.1038/nature11119

Varshney RK, Graner A, Sorrells ME (2005) Genic microsatellite markers in plants: features and applications. Trends Biotechnol 23:48–55

Vos P, Hogers R, Bleeker M et al (1995) AFLP: a new technique for DNA fingerprinting. Nucleic Acids Res 23:4407–4414. https://doi.org/10.1093/nar/23.21.4407

Vuylsteke M, Peleman JD, van Eijk MJT (2007) AFLP technology for DNA fingerprinting. Nat Protoc 2:1387–1398. https://doi.org/10.1038/nprot.2007.175

Wang Y, Xu J, Deng D et al (2016) A comprehensive meta-analysis of plant morphology, yield, stay-green, and virus disease resistance QTL in maize (Zea mays L.). Planta 243:459–471. https://doi.org/10.1007/s00425-015-2419-9

Williams RC (1989) Restriction fragment length polymorphism (RFLP). Am J Phys Anthropol 32:159–184. https://doi.org/10.1002/ajpa.1330320508

Williams JGK, Kubelik AR, Livak KJ et al (1990) DNA polymorphisms amplified by arbitrary primers are useful as genetic markers. Nucleic Acids Res 18:6531–6535. https://doi.org/10.1093/nar/18.22.6531

Xiong J-S, Ding J, Li Y (2015) Genome-editing technologies and their potential application in horticultural crop breeding. Hortic Res 2:15019. https://doi.org/10.1038/hortres.2015.19

Yan J, Yang X, Shah T et al (2010) High-throughput SNP genotyping with the Goldengate assay in maize. Mol Breed 25:441–451. https://doi.org/10.1007/s11032-009-9343-2