Gillani SFA *et al.* (2023)
Notulae Botanicae Horti Agrobotanici Cluj-Napoca
Volume 51, Issue 3, Article number 12659
DOI:10.15835/nbha51312659
Review Article

Gene mining and functional analysis related to maize (*Zea mays* L.) seed size

Syed F.A. GILLANI[1,2], Adnan RASHEED[3], Yasir MAJEED[1], Shamshad S. BUKHARI[4], Huma TARIQ[5], Shafiq-Ur-REHMAN[5], Zhan-Wu GAO[6*]

[1]Gansu Agricultural University, College of Agronomy, Lanzhou 730070, China; faheemgillani33@gmail.com; yasirmajeed5453@gmail.com
[2]Gansu Provincial Key Lab of Arid Land Crop Science, Lanzhou 730070, China
[3]Jiangxi Agricultural University, Ministry of Education/ College of Agronomy, Key Laboratory of Crop Physiology, Ecology and Genetic Breeding, Nanchang 330045, China; adnanbreeder@yahoo.com
[4]Nuclear Atomic Energy Institute of Agriculture, Tandojam Sindh, Pakistan; sayedashamshad305@outlook.com
[5]University of Poonch Rawalakot Ajk Pakistan, Laboratory of Plant Breeding and Molecular Genetics, Azad, Kashmir, Pakistan; huma.tariq006@hotmail.com; shafiqurrehman@upr.edu.pk
*[6]Baicheng Normal University, Tourism and Geographical Science Institute, Baicheng 137000, China; gaozw261@nenu.edu.cn (*corresponding author)*

Abstract

Maize has widely been studied as a model of plant-growth promoting rhizobacteria (PGPR). Here, the genome sequences of 9P. The strains, together with 26 other sequenced Maize were comparatively studied. Phylogenetic analysis of the concatenated 244 single-copy core genes suggests that the 9P. The strains and 5 other *Paenibacillus* spp., isolated from diverse geographic regions and ecological niches, formed a closely related clade (here it is called Poly-clade). Analysis of single nucleotide polymorphisms (SNPs) reveals local diversification of the 14 Poly-clade genomes. SNPs were not evenly distributed throughout the 14 genomes and the regions with high SNP density contain the genes related to secondary metabolism, including genes coding for polyketide. Recombination played an important role in the genetic diversity of this clade, although the rate of recombination was clearly lower than mutation. The distinction among people and different creatures can be gotten by relative examinations. This study reveals that both maize and its closely related species have plant growth promoting traits and they have great potential uses in agriculture and horticulture as PGPR.

Keywords: functional analysis; gene mining; maize; polymorphisms; Rhizobacteria

Introduction

Maize is the most conveyed grain crop all around. Its pack end uses and the effortlessness of advancement over contrasted biological and soil conditions has made it an appealing harvest across the world. Despite human

Received: 12 Jul 2022. Received in revised form: 12 Aug 2023. Accepted: 18 Aug 2023. Published online: 31 Jul 2023.
From Volume 49, Issue 1, 2021, Notulae Botanicae Horti Agrobotanici Cluj-Napoca journal uses article numbers in place of the traditional method of continuous pagination through the volume. The journal will continue to appear quarterly, as before, with four annual numbers.

usage, it is used as feed for trained creatures, rough materials for engineered and food endeavors and as biofuel. To also improve its agronomical characteristics, analysts have interminably endeavored to modify its genome through innate procedures. For the most part, maize characteristics were changed or adjusted through enlightenment and substance mutagens. These techniques could introduce changes in the plant genome during transparency and deoxyribonucleic destructive (DNA) fix measures. Change raising is all things considered not accurate. It can provoke both positive and antagonistic outcomes with no control over spaces of the genome to be changed (Gazzafi *et al*., 2017).

Transposon naming is another occasionally used strategy in maize genetic characteristics, whereby unequivocal transposons are used to cause changes and thus permit quality disclosure. This system is both drawn-out and can be exorbitant. It moreover prompts unpredictable changes and is massive to perform for colossal screens. Notwithstanding the way that there are many 'races' of maize, most of the fiscally grown crossbreeds transmitted two or three huge races. For the inspirations driving discussing the business meaning of maize, maize types can be parceled into four groupings not related to race. Engraving maize is the fundamental kind of maize filled in the US Corn Belt similarly as in Europe, South Africa, and China. Gouge varieties have been changed through hybridization and assurance to give a wide extent of agronomic and touch characteristics. Exceptional scratch varieties have been conveyed with novel starch characteristics. High-amylose (straight starch) and waxy (fanned starch) maize genotypes have been grown mechanically for quite a while (Rosen *et al*., 2016).

Maize plants contain both male and female regenerative developments and copy by both cross-treatment and self-preparation. In most modernly appropriate maize genotypes, the female plan projects outward from a central tail, while the male development projects out the most noteworthy mark of the tail. Residue from the tuft is passed on by the breeze to other maize plants, where treatment of the individual pieces on the ear occurs. The ears of maize may go in size from about 2.5 to in excess of 45 cm long (Wanger *et al*., 2016). The piece size, shape, and concealing similarly shift extensively. Stone maize is innately extraordinary in legacy from gouge maize and is depicted by hard, round bits. Rock maize endosperms involve overwhelmingly of hard or smooth endosperm. The agronomic characteristics of stone maize contrast from gouge maize, fundamentally due to the unique necessities of the chronicled creating regions. Popcorn is a stone kind maize that has been innately decided for its ability to expand or 'pop' when warmed. Popping happens when the pieces are immediately warmed to ~240 °C. The thick endosperm limits water rage scattering that makes pressure work inside the part until it explodes (Binder *et al.,* 2015).

Gene Mining

Plants are fixed troopers that are attached to one spot, they can't pursue supplements or escape from herbivores and microorganisms. A portion of these mixtures have been found to battle human dangers also, and cultivators have been scouring these ranges of optional metabolites for their wellbeing advancing properties for quite a long time. Current medication additionally consolidates plant compounds. Around 80% of the total populace as of now depends on ethnobotanical cures and plant drugs, for example, the antineoplastic Taxol, the antimalarial artemisinin, the pain-relieving codeine, the antidiabetic allicin, and the cardiovascular depressant quinidine. The significant expense of new medications, unpalatable results and microbial obstruction are driving a steady and recharged public interest on other option and correlative medication (Qi *et al*., 2013). Proper mechanism of the process of Gene Mining of Maize given in Figure 1.

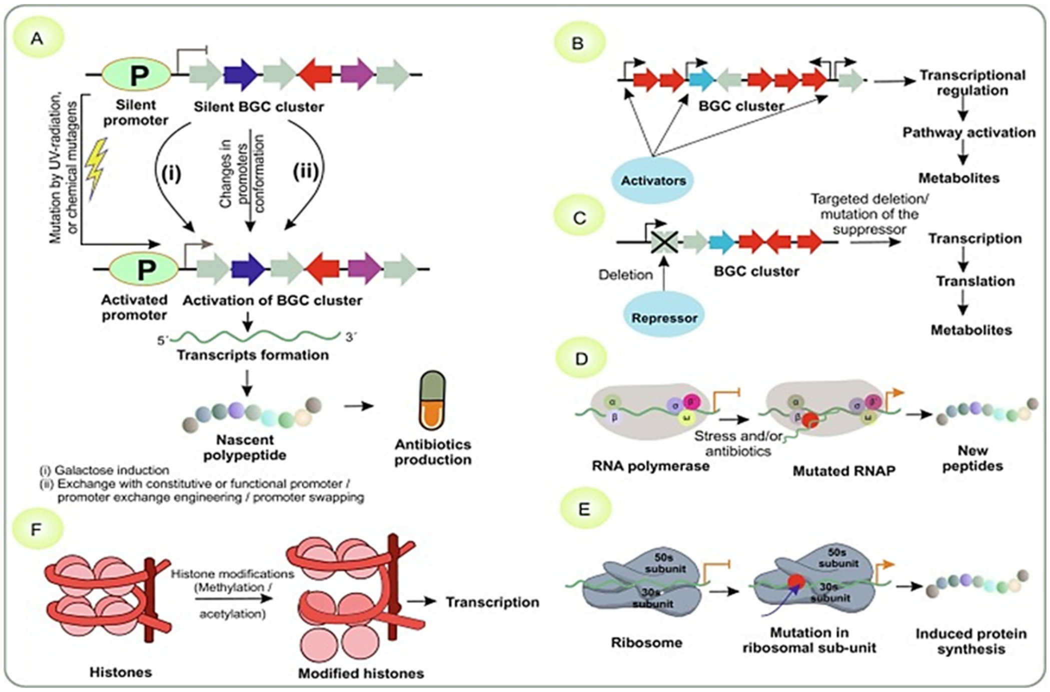

Figure 1. Proper mechanism of the process of gene mining of maize (Ayers *et al*., 1997)

Luckily, sequencing innovation has developed to furnish scientists with the devices to handle virtually these inquiries. Single Molecule, Real-Time (SMRT) Sequencing, which works like a goliath magnifying lens that can in a real sense 'see' DNA union progressively, empowers specialists to gather exceptionally coterminous and precise megabase-size extends, or contigs, of plant genomes. These 'long peruses' catch undetected primary varieties, completely unblemished qualities and administrative areas inserted in complex constructions that divided draft genomes regularly miss (Ahmadi *et al*., 2011).

Most genome-wide information is acquired at the degree of quality articulation (i.e., varieties in mRNA amount). It is regularly accepted that every individual quality translates indistinguishable RNA atoms. However, as a general rule, one quality may deliver a few distinctive isoforms by the utilization of elective advertisers, exons and eliminators. During record, elective RNA particles (i.e., isoforms) are frequently created. They can change long and vary uniquely in capacity and articulation design. On the other hand, grafted various record isoforms can significantly expand the protein-coding capability of the genome. Also, grafted isoforms deciphered from a similar quality can have essentially extraordinary and surprisingly hostile impacts (Alexa *et al*., 2010).

In that capacity, precisely catching isoform action can be critical to understanding quality construction, administrative components and coding areas. Also, covering the whole length of cDNA groupings and records can even empower the revelation of new qualities. Enter the isoform grouping (Iso-Seq) strategy, which uses since quite a while ago read innovation and requires no get together, making it an inexorably famous apparatus – particularly without reference genomes, which is a reality for some, scientists chipping away at non-model creatures and plants with genomes that are enormous and complex (Bazo *et al*., 2018).

To address these inconveniences, the Toronto/NY group sequenced the two cannabis assortments utilizing SMRT Sequencing which gave new experiences into the plan of the chromosomes and the cannabinoid biosynthetic qualities, including revelation of considerable adjustment and quality duplications at the intently connected THC and CBD corrosive synthase quality loci. Maybe than resolve the THCAS/CBDAS secret, nonetheless, the hereditary guide brought up more issues. "They are not isoforms at a generally comparable locus, and no likeness THCAS (deactivated or not) is found in hemp", the creators composed. Their perceptions recommended that either polymorphisms or differential guideline of sweet-smelling prenyltransferase (AP) adds to cannabinoid creation, apparently by controlling substrate fixation for THCAS and CBDAS (Kang *et al*., 2015).

Purple Kush has more noteworthy than five-overlay higher record levels of AP than Finola, with no distinction in duplicate number, recommending that AP protein levels might be higher in drug-type plants halfway because of contrasts in record levels. One of home-grown genomics' greatest examples of overcoming adversity additionally represents the limits that accompany deficient genomic and transcriptomic inclusion. Chinese researcher Youyou Tu got a Nobel Prize in Physiology or Medicine in 2015 for her disclosure of the counter jungle fever capacity of artemisinin, an endoperoxide sesquiterpene lactone disengaged from sweet wormwood a yearly spice of the Asteraceae family. Artemisinin-based mix treatments (ACTs), suggested by the World Health Organization for the treatment of simple intestinal sickness brought about by the Plasmodium falciparum parasite, have saved large number of lives. Other helpful impacts have likewise been accounted for artemisinin for infections like malignant growth, tuberculosis and diabetes, so request is high for the compound. Be that as it may, plant-based creation is battling to satisfy the worldwide need because of the low measure of artemisinin delivered in A. annua leaves (0.1%-1.0% of dry weight) (Terole *et al*., 2015).

Extensive transcriptome sequencing permitted the Choi group to conquer this obstruction. They had the option to distinguish 76 cytochrome P450s and arrange them into remarkable families, which ought to empower their individual capacities to be anticipated with exactness. They were likewise ready to bits together another vital riddle: anti-toxin biosynthesis. This is really where they tracked down the most noteworthy number of isoforms (1,250) and chemicals. Suspected to effect neurodegenerative issues, cardiovascular infections, diabetes and malignant growth, Panax ginseng and Panax notoginseng contain novel saponins called ginsenosides. Study into these glycosylated triterpenes has been hampered, notwithstanding, because of the lethargic development (~4 years/age), long age time, low seed creation and convoluted genome construction of Panax plants (Fu *et al*., 2012). Basic Mechanism of how the genetic changes occur at different locus position shows in Figure 2.

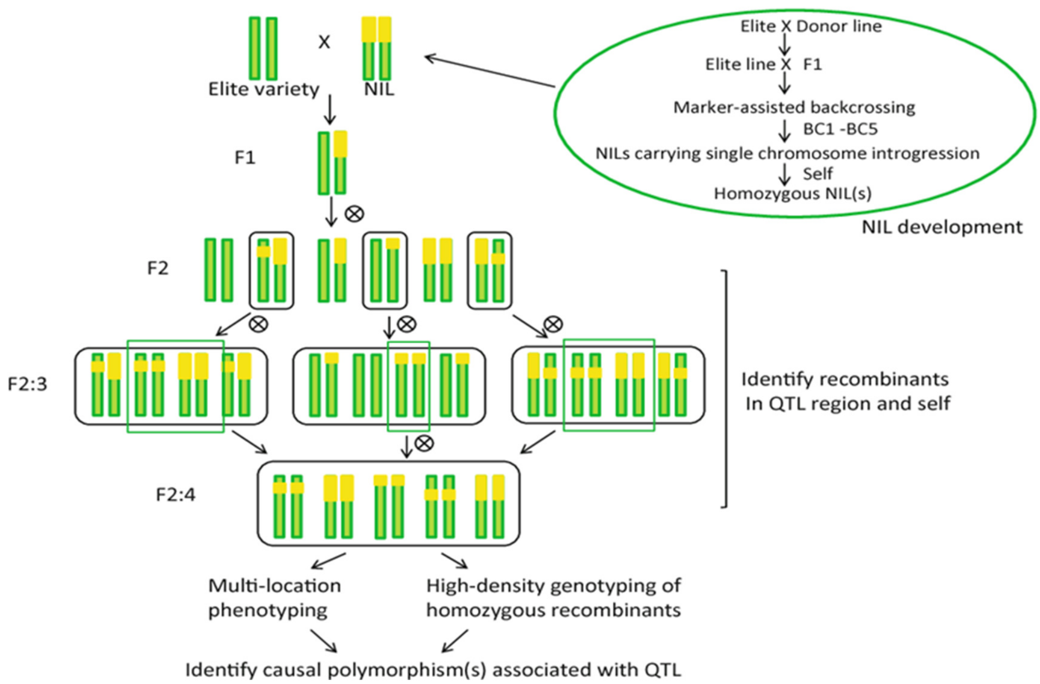

Figure 2. Basic Mechanism of how the genetic changes occur at different locus position (Laurent *et al*., 2000)

The principal all over again get together of a Panax genome – a 2.36 Gbp diploid P. notoginseng with 35,451 protein-encoding qualities – was at long last detailed as a pre-print in July 2018 (5) by a group from the Chinese University of Macau drove by Simon Ming-Yuen Lee. An all over again gathering of a 2.98 Gbp genome (with 59,352 commented on qualities) of the tetraploid P. ginseng cultivar Chunpoong (ChP), created by a group from Seoul National University drove by Tae Jin Yang, followed presently. Sequencing of both DNA and mRNA empowered scientists to bring profound jumps into the ginsenoside biosynthetic hardware, yet in addition its guideline and metabolic use. On account of *P. ginseng*, Yang *et al.* (2013) developed genome-scale metabolic organizations covering almost 5,000 quality items, catalyzing 2,194 responses and 2,003 exceptional metabolites (Li *et al.*, 2020).

Ginsenosides gather contrastingly in roots, leaves, stems, blossom buds and berries, in amounts changing with tissue, age, climate and cultivar. Yang's group had the option to decide from whole genome sequencing that the high ginsenoside substance in more established P. ginseng roots are likely the consequence of transportation from shoot tissues instead of dynamic biosynthesis. Co-articulation examination utilizing RNA sequencing information recognized significant chemicals with which ginsenoside creation co-developed. On account of *P. notoginseng*, two sorts of ginsenosides (PPD and PPT) with contradicting organic exercises (supportive of angiogenesis and against angiogenesis) can be found in a similar plant. Simply by completely portraying the whole genome of the plant, just as isoforms from eight of its constituent parts (Du *et al.*, 2014) was the Lee group ready to verify that the flying parts (e.g., leaf and bloom) contain a higher bounty of PPD contrasted with roots. A long way from being a relic of an old-fashioned past, therapeutic plants and natural cures have educated a lot regarding present day medication and could contribute an extraordinary arrangement to sound, science-based arrangements of things to come (Gao *et al.*, 2013).

We actually have a lot to find out about the hereditary and epigenetic components of these possibly wellbeing advancing plants. Fortunately, current sequencing stages empower us to examine the extraordinary primary association of qualities and the administrative systems fundamental (Cannon *et al.*, 2012) their demeanor designs, permitting the age of inventories of specific digestion in manners inconceivable to the cultivators who initially saddled their recuperating properties hundreds of years prior. As Tessa Moses and Alain Goossens bring up in the Journal of Experimental Botany, all living plant species on the planet together add to a more prominent synthetic variety of bioactive mixtures than any man-made substance library (Garvin *et al.*, 2018). The study was found to be very useful and important for future research.

Functional genome analysis

Though human genomes are about 99.9% indistinguishable, the excess 0.1% is the explanation of distinction between individuals brought about by various variations. Since 2003, the total grouping of human genome, its explanation and expanded headway of sequencing innovations (I.e., Sanger and Next-age sequencing; NGS) have given every one of the important conditions to the recognizable proof of all variations in human coding and non-coding grouping. Albeit the strategy for variation recognition is presently turning into an everyday practice, the critical inquiry all through numerous years concerns the capacity of distinguished variations. The asset of significant data about utilitarian genomics are a few enormous scope projects, (Murrey *et al.*, 2012) for example, the ENCODE project, the principle objective of which was to distinguish every one of the useful components, remembering administrative components for both coding and non-coding regions. As per another, the 1000 Genomes Project, there are around 20,000–23,000 variations in interchangeable and no synonymous areas of the human genome. Despite the fact that not every one of them are practically significant, 530–610 of the variations have utilitarian effect by causing in frame cancellations and inclusions, untimely stop codons, frame shifts, or by disturbing join destinations (Gupta *et al.*, 2020).

In spite of various investigations, researchers are as yet confronting a tremendous test in disentangling what the succession implies and in choosing whether or not a discovered variation is pathogenic. A pathogenic variation can prompt infection or cause various issues. Be that as it may, comprehension of pathogenic components sets out a freedom to forestall extreme outcomes by creating novel demonstrative apparatuses and by planning profoundly successful medicines for the infection. To accomplish this point it is important to perform huge scope utilitarian genome examination that includes various fields of study: genomics, epigenomics, transcriptomics, proteomics, and interactomes. To depict the elements of qualities and proteins just as to investigate the connection between the genotype and the aggregate, an enormous number of different strategies, including model frameworks (e. g., CRISPR-Cas9), can be utilized (Hencock et al., 2020). Different tests performed on number of variants of plants showing NR sequences and the redundancy mentioned in Table 1.

Table 1. Different tests performed on number of variants of plants showing NR sequences and the redundancy

Sources of ESTs	Number of SSR-ESTs	Consensus sequences	Singletons	Total number of NR sequences	% Reduction in redundancy
Barley	3462	269	600	869	75
Wheat	3624	335	302	637	83
ITEC	2277	230	457	687	70
Maize	4107	408	385	793	81
Sorghum	3976	430	498	928	77
Rice	7160	752	1001	1753	76
Total	24606	2424	3243	5667	77

The critical way to effective arrangement examination is to adjust the succession important to another grouping whose capacity is referred to (normally name as the reference genome). It very well may be valuable when the quality capacity is obscure however is developmental identified with another quality whose capacity is characterized. In such a case, it very well may be thought that the obscure quality has something very similar or comparable capacity. Moreover, the arrangements may be checked to track down the huge matches (Martin *et al.*, 2011) between the segments of a grouping that have been recently portrayed as enormously affecting the genomics work. To analyze the information, it is important to look for data in changed biomedical data sets. Perhaps the greatest wellspring of biomedical and genomic data is the NCBI (National Center for Biotechnology Information), which gives admittance to different data sets like PubMed, Entrez Gene, OMIM, Variation Viewer, dbSNP, and others (Ince *et al.*, 2014). Pictorial representation of Genomic Functional analysis shows in Figure 3.

There are some of the types that are as follows;

- Epigenomics
- Transcriptomics
- Proteomics and Interactomics
- Functional Genomics Integrating Model Systems

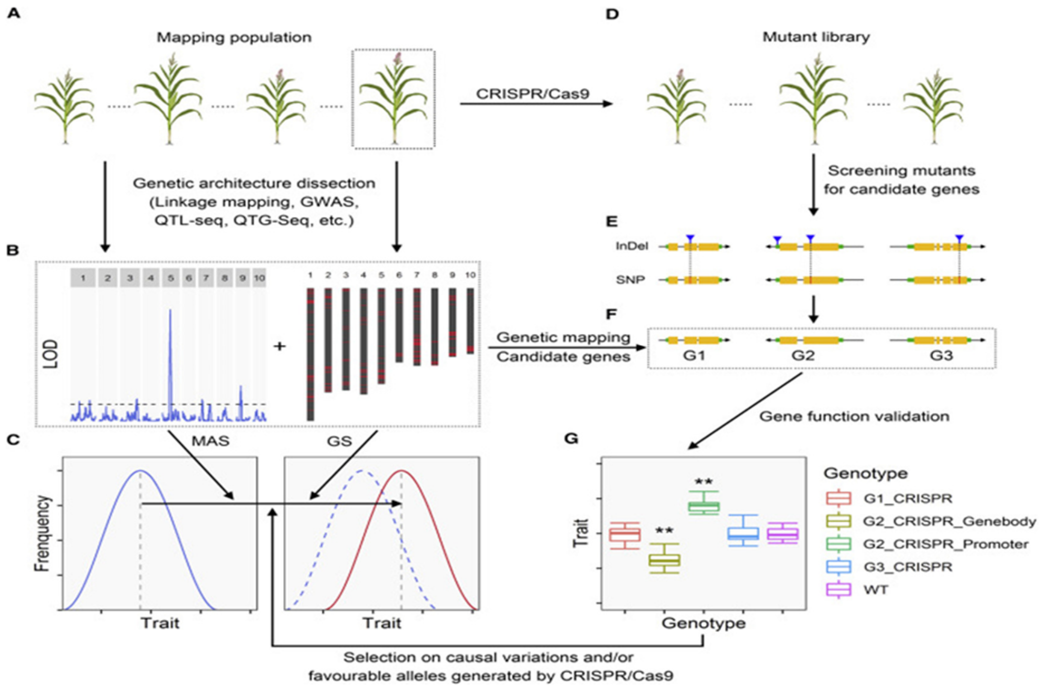

Figure 3. Pictorial representation of genomic functional analysis (Bowen *et al*., 2011)

Epigenomes

For utilitarian examination, it is critical to consider epigenetic adjustments like DNA methylation and histone alterations, since they influence quality articulation with no progressions in the basic DNA succession. A segment of these combinations has been found to fight human perils additionally, and cultivators have been scouring these scopes of discretionary metabolites (Reddy *et al*., 2012) for their prosperity propelling properties for a serious long time. Current prescription moreover combines plant compounds. Around 80% of the all-out people as of now relies upon ethnobotanical fixes and plant drugs, for instance, the antineoplastic Taxol, the antimalarial artemisinin, the torment diminishing codeine, the antidiabetic allicin, and the cardiovascular depressant quinidine. The extended sensibility and intricacy of genetic sequencing development is making the total of this possible. To fathom the full metabolic capacity of plants, broad genomic information ought to be gotten together with transcriptomic, proteomic and metabolomics data (Onus *et al*., 2010).

We ought to have the alternative to react to questions, for instance, the quality duplication, similar to whole genome duplications (WGDs) and neighborhood (couple) duplication (LDs), can in like manner accept a critical part specifically processing, including the outpouring of flavonoid related characteristics. Fortunately, sequencing advancement has created to outfit researchers with the gadgets to deal with practically these requests. Single Molecule, Real-Time (SMRT) Sequencing, which works like a goliath amplifying focal point that can from a genuine perspective 'see' DNA association logically (Reddy *et al*., 2012), enables experts to assemble particularly coterminous and exact mega base-size expands, or contigs, of plant genomes. These 'long scrutinizes' catch undetected essential assortments, totally unsullied characteristics and regulatory regions embedded in complex developments that separated draft genomes routinely miss (Burger *et al*., 2019). Most genome-wide data is gained at the level of value verbalization (i.e., assortments in mRNA sum). It is consistently acknowledged that each individual quality deciphers vague RNA iotas. Be that as it may, when in doubt, one quality may convey a couple unmistakable isoforms by the usage of elective sponsors, exons and eliminators. During record, elective RNA particles (ie, isoforms) are every now and again made. They can change long and fluctuate exceptionally in limit and explanation plan. Then again, joined different record isoforms can altogether grow the protein-coding capacity of the genome. Additionally, united isoforms translated from a comparable quality can have basically exceptional and shockingly unfriendly effects (Komada *et al*., 2003). Epigenomics in maize and detailed mechanism of its working showed in Figure 4.

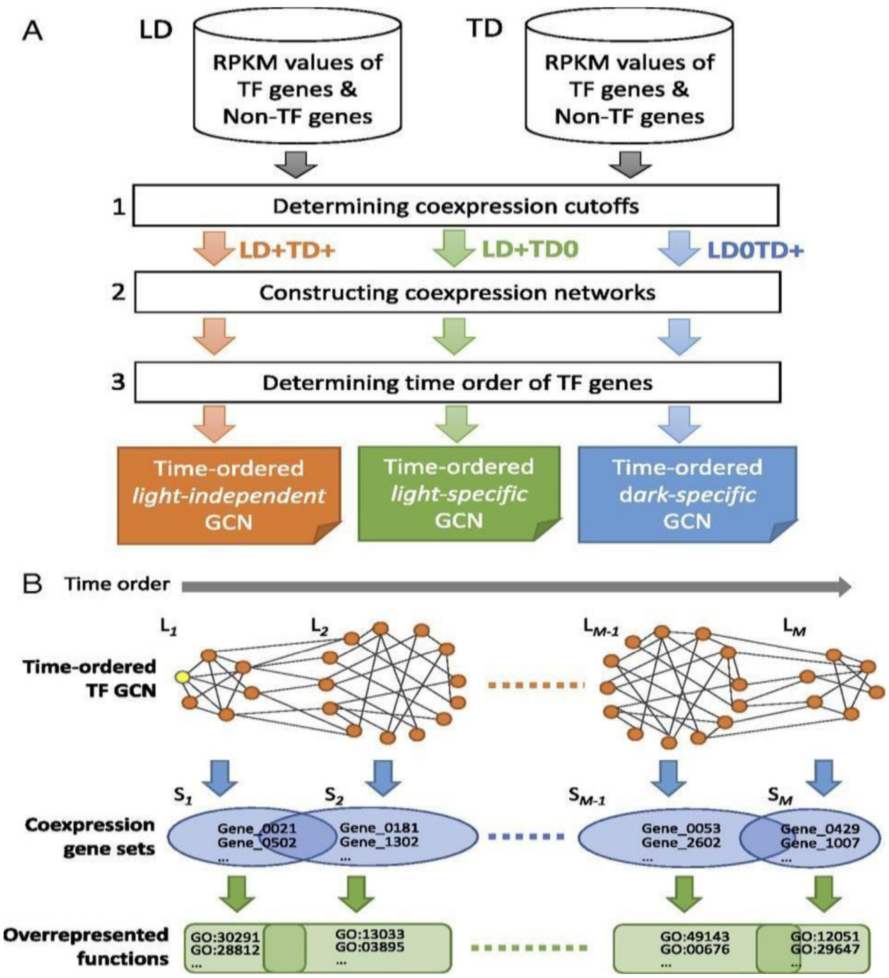

Figure 4. Epigenomics in maize and detailed mechanism of its working (Russell *et al*., 2013)

Transcriptomics

Basically, by totally depicting the entire genome of the plant, similarly as isoforms from eight of its constituent parts, was the Lee gathering prepared to check that the flying parts (e.g., leaf and blossom) contain a higher abundance of PPD appeared differently in relation to roots. They perceived a couple of key characteristics, including a couple seen strangely, similarly as their record factor limiting areas and other related parts drew in with the ginsenosides association pathway. As Yang raises, such information will be fundamental to engaging in silico metabolic planning to anticipate contender characteristics related with overproduction of liked metabolites (Zhang *et al*., 2010) and in as such accelerate in everyday metabolic planning cycles. Far from being a relic of an older style past, restorative plants and regular fixes have taught a great deal with respect to introduce day medicine (Jia *et al*., 2017) and could contribute an exceptional course of action to sound, science-based plans of what might be on the horizon (Dutta *et al*., 2011).

We really have a great deal to get some answers concerning the innate and epigenetic parts of these conceivably prosperity propelling plants. Luckily, current sequencing stages engage us to inspect the uncommon essential relationship of characteristics and the regulatory frameworks major their disposition plans, allowing the period of inventories of explicit assimilation in habits unfathomable to the cultivators (Grener *et al*., 2011) who at first outfitted their recovering properties many years earlier. By returning to our basic establishments through roots, and getting old discernments together with present day nuclear mining, we can announce another time of recovering and medicine divulgence (Jones *et al*., 2002).

Gigantic extension genome projects have fundamentally changed the essence of science. Genomics has habitually been suggested as another field that has provoked an adjustment of viewpoint in the way in which science is performed. Meanwhile the post-genomic period has emerged by misusing the tremendous (Stein *et al*., 2012) proportion of genome gathering data. It has gotten possible to look at science in a substitute way as in nowadays researchers don't need to advance toward natural requests in a theory driven way anyway rather can accumulate and examine data in a more non-uneven and more broad style. New sorts of consistent requests can be presented and new kinds of examinations (Bowen *et al*., 2011) can be performed at a remarkable speed. Continuous inventive advances and the speedy improvement of novel gadgets as of now award the questioning of an all-out genome simultaneously and in a singular preliminary (Levin *et al*., 2001). Transcriptomics analysis of maize at different genomic level shows in Figure 5.

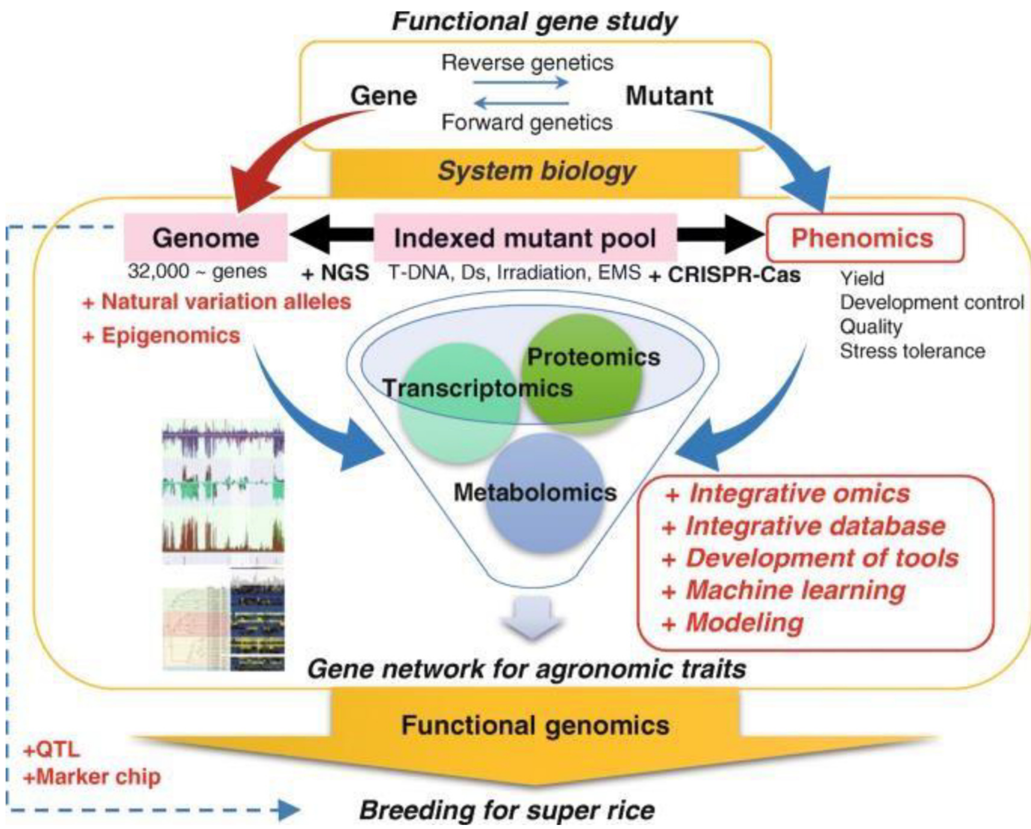

Figure 5. Pictorial diagram of functional gene study in maize (Schmitz *et al*., 2012)

Knowing the particular progression besides, space of the large number of characteristics of a given animal is figuratively speaking the underlying move towards perceiving how all of the bits of a natural structure collaborate. In such manner reasonable genomics is the imperative (Zhang *et al.*, 2019) method to manage changing sum into quality. Valuable genomics is a general technique toward perceiving how the characteristics of a natural element cooperate by assigning new abilities to cloud characteristics. Information about the speculated limit of a dark quality may be closed (Herr *et al*., 2014) from its gathering structure using unquestionably alluded to components of similar characteristics as the justification assessment (Province *et al*., 2001).

Functional Genomics Integrating Model Systems

From the practical perspective, examination of proteomics and interactomes is pretty much as crucially significant as recently portrayed investigation (Noda 2014) of genomics, epigenomics, and transcriptomics, in light of the fact that a few examinations show that quality articulation at DNA or mRNA levels is generously unaltered, in spite of the fact that it influences the protein capacity and the other way around. Proteins play out a huge range of capacities (Feng *et al.,* 2014) inside life forms, however unusual protein articulation that happens because of post-transcriptional changes or protein communication with another protein or nucleic acids upsets cell work (Aliferis *et al*., 2003)

Contingent upon the plan of the test, there are two notable procedures for protein measurement: immunoassays or neutralizer free discovery techniques. Immunoassay, for example, the catalyst connected immunosorbent examine (ELISA), is a generally utilized technique (Muller *et al*., 2016) because of its high affectability and solid particularity. In any case, it is only the new variety of these techniques to the simultaneous assessment of tremendous amounts of characteristics, proteins (Keller *et al*., 2016), substance constituents and oddity plants that has allowed the creation of the current down to earth genomic development stage. A couple of methodologies have viably become standard devices for inspecting quality limit while others are at this point in their beginning phases (Martin *et al*., 2014). The rapidly emerging field of close to genomics has yielded enthusiastic results. Close to genome examination has gotten conceivable with the openness of different completely sequenced genomes (Wolf *et al*., 2013). Genomic analysis performed on different variants showed results of different combinations of nitrogenous bases in Table 2. Assessment of complete genomes (Mock *et al*., 2014) between natural elements consider overall points of view on genome progression and the availability of various completely sequenced genomes fabricates the judicious power in deciphering the privileged information in genome setup, limit and improvement. In like manner, assessment of human characteristics with characteristics (Freeling 2020) from various genomes in a genomic scene could help apportion novel capacities with respect to un-clarified characteristics. Here, we talk about the actually used strategies for relative genomics and their derived allowances in genome biology. As on Jan 25, 2007, 472 genomes are completely sequenced (Bird *et al*., 2011) yet then another 498 are in progress. The quick headway in genome sequencing demands more relative examination to gain new encounters into extraordinary, biochemical, inherited, metabolic, and physiological pathways (Ogata *et al*., 2014).

In each genome, and checked districts kept up in both prokaryotic and eukaryotic social occasions of living creatures. Close to genomics not solely can follow out the formative association between life frames yet furthermore differentiations and likenesses inside (Sanbor *et al.,* 2014) and between species. The differentiation among individuals and various animals can be gotten by relative assessments. To report the obvious features of individuals, the most enlightening investigation incorporates standing out individuals from our closest relatives, the chimpanzees (Kim *et al*., 2012) and apes. Genome correspondence the procedure for choosing the correct correspondence of chromosomal segments and valuable parts across the species contemplated is the underlying stage in relative genomics (Paz *et al*., 2017).

This incorporates choosing orthologous (characteristics veered after a speciation event) parts of DNA that dive from comparable region in the fundamental begetter (Zhu *et al*., 2013) of the species took a gander at, and paralogous (characteristics isolated after a duplication event) regions that arose by duplication events going before the uniqueness of the species ponder. The arranging of areas across two genomes can be facilitated without duplication events; one-to many if a region has gone through duplication or incident in one of the creature classes (Jin *et al*., 2015) or many-to-various if duplication/hardship has occurred in the two heredities. Fitch et al., cultivated a method called BBH (Best Bidirectional Hits), which recognizes quality consolidates that are best matches of each other as orthologous (Twell *et al.*, 2015). Pictorial diagram of functional gene study in maize presented in Figure 6.

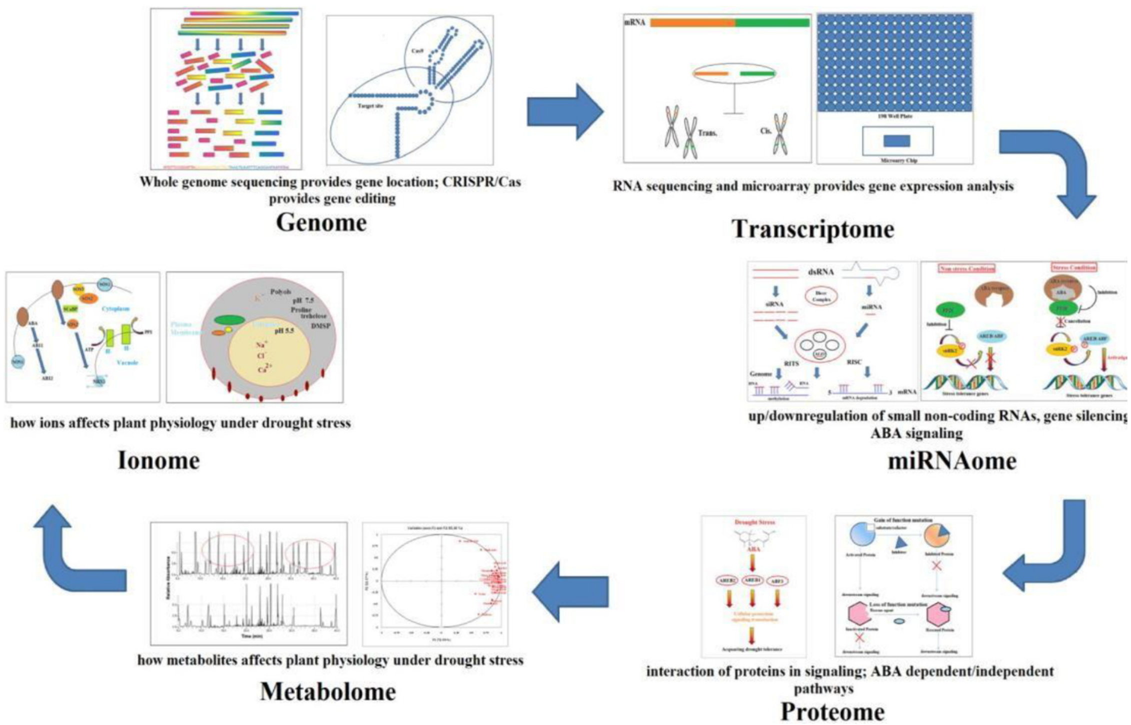

Figure 6. Different levels of plant functional genomics. Adapted from Zhang *et al.* (2012).

Table 2. Genomic analysis performed on different variants showed results of different combinations of nitrogenous bases

SSR Motif	Barely	Wheat	ITEC	Maize	Sorghum	Rice	Total
GA/CT	244	81	112	198	147	373	1155
CA/GT	37	31	24	40	70	24	226
AT/TA	14	13	8	51	52	5	2036
GC/CG	9	0	10	2	0	1	22
AGC/TGC	106	125	84	109	199	195	818
AAC/TTC	7	7	3	21	16	25	97
AAC/TTG	14	13	11	13	24	22	426
AAT/TTA	44	65	27	24	40	226	662
GGC/CGG	33	402	171	21	14	21	78
CCT/GGA	5	3	1	42	7	20	2377
Total	845	1066	651	721	1196	2502	6981

Understanding the family line of the reasonable segments stood out is central from our game plan and employments of genome assessment (Sano *et al.*, 2014). Most close to methods have focused in on adjusted orthologous areas, anyway it is correspondingly basic to see what segments have gone through duplication events, and which sections were lost since the divergence (Tran *et al.*, 2013) of the species. Taking a gander at divides that arose before the uniqueness of the species may achieve some unsatisfactory understandings of collection protection and contrast. Further, inside seeing quality duplication, a bit of the groundbreaking necessities (Pilu *et al.*, 2013) that a region is under are soothed, and uniform models of headway now don't get the essential decision for these regions. Thusly, our techniques for choosing quality correspondence ought to address duplication and adversity events, and assurance that the parts we consider are orthologous (Kun *et al.,* 2016).

Plant functional genomics

Enormous scope genome projects have significantly changed the face of science. Genomics has frequently been alluded to as a new field that has prompted (Kranz *et al*., 2015) a change in outlook in the manner in which science is performed. In the interim the post-genomic period has arisen by exploiting the huge measure of genome grouping information (Hansen *et al*., 2017). It has gotten conceivable to take a gander at science in an alternate manner as in these days scientists don't have to move toward organic inquiries in a speculation driven path however rather can gather and investigate information (Smolen *et al*., 2015) in a more non-one-sided and more extensive style. New kinds of logical inquiries can be posed and new sorts of analyses can be performed at an exceptional speed. At present the mass of genome information is being changed over into gene function information, implying that worth is added to the nucleotide grouping assortments (Cai *et al*., 2012).

Nonetheless, to nail down the specific capacity of obscure qualities it is important to see every quality's job in the unpredictable arrangement of all quality exercises in the plant cell (Willioms *et al*., 2017). Quality capacity examination in this manner requires the investigation of transient and spatial quality articulation designs. The most decisive data about changes in quality articulation levels can be acquired from investigation of the shifting subjective and quantitative changes of courier RNAs, proteins and metabolites (Van *et al*., 2018). New advances have been created to permit quick and profoundly equal estimations of these constituents of the cell that make up quality action. The essential standards basic the distinctive insightful advances have been known for a very little time (Tohge *et al*., 2015a).

Conclusions

After being domesticated from teosinte, cultivated maize (*Zea mays* ssp. *mays*) spread worldwide and now is one of the most important staple crops. Here, we provide an overview of the history of maize domestication and key genes controlling major domestication-related traits, review the currently available resources for functional genomics studies in maize, and discuss the functions of most of the maize genes that have been positionally cloned and can be used for crop improvement. Finally, we provide some perspectives on future directions regarding functional genomics research and the breeding of maize and other crops. We have defined a family of 33 maize Pap genes which we predict, on the basis of transcript accumulation and similarity to proteins characterized in other plants, to be functionally diverse and to play a role in both the P deprivation response and more generally in maize stress responses and development. From what is known of the post-transcriptional regulation of PAPs in other plants, it is probable that the capacity for functional divergence is even greater than revealed by this first characterization. Furthermore, groups of closely related sequences present dramatically different patterns of transcript accumulation, illustrating a capacity for rapid adoption of new biological roles during the radiation of the maize Pap family. Ultimately, a more complete understanding of the roles of individual maize PAPs will require functional and biochemical analysis. Given the availability of a reference genome and the increasing availability of resources for reverse genetics, it is now feasible to conduct a functional genomics analysis of a maize gene family. On the basis of the characterization presented here, we have selected a number of candidate genes and initiated a program of reverse screening. We anticipate that functional characterization of maize PAPs will facilitate their use as direct targets of selection and manipulation, as well as valuable reporters of plant nutrient and stress status.

Authors' Contributions

SFAG prepared the draft, AR reviewed and improved the paper. YM, SSB, HT, SR provided technical assistance. ZWG supervised the study.

All authors read and approved the final manuscript.

Ethical approval (for researches involving animals or humans)

Not applicable.

Acknowledgements

National Natural Science Foundation of China (No.3177); Key Project of Science and Technology Research in the 13th Five-Year Plan of Education Department of Jilin Province (No.41 of Jijiao Kehe, 2016); Talents of Jilin Province Supported Project (2020047).

Conflict of Interests

The authors declare that there are no conflicts of interest related to this article.

References

Achard P, Herr A, Baulcombe DC, Harberd NP (2014). Modulation of floral development by a gibberellin-regulated micro-RNA. Development 131:3357-3365. *https://doi.org/10.1242/dev.01206*

Addo-Quaye C, Eshoo TW, Bartel DP, Axtell MJ (2018). Endogenous siRNA and miRNA targets identified by sequencing of the Arabidopsis degradome. Current Biology 18:758-762. *https://doi.org/10.1016/j.cub.2008.04.042*

Ahmadi J, Fotokian MH (2011). Identification and mapping of quantitative trait loci associated with salinity tolerance in rice (*Oryza sativa*) using SSR markers. Iran Journal of Biotechnology 9:21-30. *https://doi.org/10.5897/AJB2018.16661*

Akagi T, Ikegami A, Yonemori K (2012). DkMyb2 wound-induced transcription factor of persimmon (Diospyros kaki Thunb.), contributes to proanthocyanidin regulation. Planta 232:1045-1059. *https://doi.org/10.1007/s00425-010-1241-7*

Alexa A, Rahnenfuhrer J (2010). TopGO: enrichment analysis for gene ontology. R Package Version 2:2010. *https://doi.org/10.1038/s41467-022-31280-w*

Apweiler R, Attwood TK, Bairoch A, Bateman A, Birney E (2011). The InterPro database, an integrated documentation resource for protein families, domains and functional sites. Nucleic Acids Research 29:37-40.

Aya K, Ueguchi-Tanaka M, Kondo M, Hamada K, Yano K (2019). Gibberellin modulates anther development in rice via the transcriptional regulation of GAMYB. Plant Cell 21:1453-1472. *https://doi.org/10.1105/tpc.108.062935*

Bailey TL, Gribskov M (2018). Combining evidence using p–values: application to sequence homology searches. Bioinformatics 14:48-54. *https://doi.org/10.1093*

Bailey TL, Williams N, Misleh C, Li WW (2016). MEME: discovering and analyzing DNA and protein sequence motifs. Nucleic Acids Research 34:W369-W373. *https://doi.org/10.1093/nar/gkl198*

Ballester AR, Molthoff J, de Vos R, Hekkert BL, Orzaez D (2010). Biochemical and molecular analysis of pink tomatoes: deregulated expression of the gene encoding transcription factor SlMYB12 leads to pink tomato fruit color. Plant Physiology 152:71-84. *https://doi.org/10.1104/pp.109.147322*

Balsalobre TWA., da Silva Pereira G, Margarido GRA, Gazaffi R, Barreto FZ, Anoni CO, Carneiro MS (2017). GBS-based single dosage markers for linkage and QTL mapping allow gene mining for yield-related traits in sugarcane. BMC genomics 18(1):1-19. *https://doi.org/10.1186/s12864-016-3383-x*

Baumann K, Perez–Rodriguez M, Bradley D, Venail J, Bailey P (2017). Control of cell and petal morphogenesis by R2R3 MYB transcription factors. Development 134:1691-1701. *https://doi.org/10.1242/dev.02836*

Bazzo BR, de Carvalho LM, Carazzolle MF, Pereira GAG, Colombo CA (2018). Development of novel EST-SSR markers in the macaúba palm (*Acrocomia aculeata*) using transcriptome sequencing and cross-species transferability in Arecaceae species. BMC Plant Biology 18:276. *https://doi.org/10.1186/s12870-018-1509-9*

Boddu J, Jiang C, Sangar V, Olson T, Peterson T (2016). Comparative structural and functional characterization of sorghum and maize duplications containing orthologous myb transcription regulators of 3-deoxyflavonoid biosynthesis. Plant Molecular Biology 60:185-199. *https://doi.org/10.1007/s11103-005-3568-1*

Busson-Le Coniat M, Salomon-Nguyen F, Hillion J, Bernard OA, Berger R (2019). MLL-AF1q fusion resulting from t(1;11) in acute leukemia. Leukemia 13:302-306. *https://doi.org/10.1038/sj.leu.2401299*

Byrne ME, Barley R, Curtis M, Arroyo JM, Dunham M (2018). Asymmetric leaves1 mediates leaf patterning and stem cell function in Arabidopsis. Nature 408:967-971. *https://doi.org/10.1038/35050091*

Cai H, Tian S, Dong H (2012). Large scale *in silico* identification of MYB family genes from wheat expressed sequence tags. Molecular Biotechnology. *https://doi.org/10.1007/s12033-011-9486-3*

Cannon SB, Shu H, Baumgarten AM, Spangler R, May G, Cook DR, Young ND (2012). Diversity, distribution, and ancient taxonomic relationships within the TIR and non-TIR NBS-LRR resistance gene subfamilies. Journal of Molecular Evolution 54:548-562. *https://doi.org/10.1007/s0023901-0057-2*

Carine AT, Wang BB, Majesta SO, Shweta D, Zhu HY, Bruce R, Nevin DY, Steven BC (2018). Identification and characterization of nucleotide-binding site-leucine-rich repeat genes in the model plant *Medicago truncatula*. Plant Physiology 146:5-21. *https://doi.org/10.1104/pp.107.104588*

Ceasar SA, Ignacimuthu S (2018). Efficient somatic embryogenesis and plant regeneration from shoot apex explants of different Indian genotypes of finger millet (*Eleusine coracana* (L.) Gaertn.). In Vitro Cell Devision Biology Plant 44:427-435. *https://doi.org/10.1007/s11627-008-9153-y*

Celenza JL, Quiel JA, Smolen GA, Merrikh H, Silvestro AR (2015). The Arabidopsis ATR1 Myb transcription factor controls indolic glucosinolate homeostasis. Plant Physiology 137:253-262. *https://doi.org/10.1104/pp.104.054395*

Chen C, Zhou P, Choi YA, Huang S, Gmitter FG Jr (2016). Mining and characterizing microsatellites from citrus ESTs. Theoretical and Applied Genetics 112:1248-1257. *https://doi.org/10.1007/s00122-006-0226-1*

Chen DM, Kang H, Liu CJ (2011). An overview on the potential quaternary glacial refugia of plants in China mainland. Bulletin of Botanical Research 31:623-632. *https://doi.org/10.7525/j.issn.1673-5102.2011.05.019*

Chen LY, Cao YN, Yuan N, Nakamura K, Wang GM, Qiu YX (2015). Characterization of transcriptome and development of novel EST-SSR makers based on next-generation sequencing technology in *Neolitsea sericea* (Lauraceae) endemic to East Asian land-bridge islands. Molecular Breeding 35:187. *https://doi.org/10.1007/s11032-015-0379-1*

Chen SY, Dong ML, Zhang Y, Qi SZ, Liu XZ, Zhang JF (2020). Development and characterization of simple sequence repeat markers for, and genetic diversity analysis of *Liquidambar formosana*. Forests 11:203. *https://doi.org.10.3390/f11020203*

Chennaveeraiah MS, Hiremath SC (2011). Cytogenetics of minor millets. In: Tsuchiya T, Gupta PK (Eds). Chromosome Engineering in Plants Genetics Breeding and Evolution. Elsevier, Amsterdam, pp 613-627. *https://doi.org/10.1016/B978-0-444-88259-2.50034-5*

Chilingaryan A, Gevorgyan N, Vardanyan A, Jones D, Szabo A (2002). Multivariate approach for selecting sets of differentially expressed genes. Math Bioscience 176:59-69. *https://doi.org/10.1016/S0025-5564(01)00105-5*

Cho SJ, Hermsmeier MA (2002). Genetic algorithm guided selection: variable selection and subset selection. Journal of Chemical Information and Computer Sciences 42:927-936. *https://doi.org/10.1021/ci010247v*

Cho YG, Ishii T, Temnykh S, Chen X, Lipovich L, McCouch SR, … Cartinhour S (2010). Diversity of microsatellites derived from genomic libraries and Gen Bank sequences in rice (*Oryza sativa).* Theoretical and Applied Genetics 100:713-722. *https://doi.org/10.1007/s001220051343*

Colquhoun TA, Kim JY, Wedde AE, Levin LA, Schmitt KC (2012). PhMYB4 fine-tunes the floral volatile signature of *Petunia* x hybrida through PhC4H. Journal of Experimental Botany 62:1133-1143. *https://doi.org/10.1093/jxb/erq342*

Cone KC, Cocciolone SM, Burr FA, Burr B (2013). Maize anthocyanin regulatory gene pl is a duplicate of c1 that functions in the plant. Plant Cell 5:1795-1805. *https://doi.org/10.1105/tpc.5.12.1795*

Conesa A, Gotz S, Garcia-Gomez JM, Terol J, Talon M, Robles M (2015). Blast2GO: a universal tool for annotation, visualization and analysis in functional genomics research. Bioinformatics 21:3674-3676. *https://doi.org/10.1093/bioinformatics/bti610*

Czemmel S, Stracke R, Weisshaar B, Cordon N, Harris NN (2019). The grapevine R2R3-MYB transcription factor VvMYBF1 regulates flavonol synthesis in developing grape berries. Plant Physiology 151:1513-1530. *https://doi.org/10.1104/pp.109.142059*

Dang ZH, Huang L, Jia YY, Lockhart PJ, Fong Y, Jan YY (2020). Identification of genic SSRs provide a perspective for studying environmental adaptation in the endemic shrub *Tetraena mongolica*. Genes 11:322. *https://doi.org/10.3390/genes11030322*

Dong, ML, Wang ZW, He QW, Zhao J, Fan ZR, Zhang, JF (2018). Development of EST-SSR markers in *Larix principis-rupprechtii* Mayr and evaluation of their polymorphism and cross-species amplification. Trees 32:1559-1571. *https://doi.org/10.1007/s00468-018-1733-9*

Dubos C, Le Gourrierec J, Baudry A, Huep G, Lanet E (2018). MYBL2 is a new regulator of flavonoid biosynthesis in *Arabidopsis thaliana*. The Plant Journal 55:940-953. *https://doi.org/10.1111/j.1365-313X.2008.03564.x*

Durand J, Catherine B, Emilie C, Frigerio JM (2010). A fast and cost-effective approach to develop and map EST-SSR markers: oak as a case study. BMC Genomics 11:570-583. *https://doi.org/10.1186/1471-2164-11-570*

Dutta S, Kumawat G, Singh BP,Gupta DK., Singh S, Dogra V (2011). Development of genic-SSR markers by deep transcriptome sequencing in pigeon pea *Cajanus cajan* (L.) Millspaugh. BMC Plant Biology 11:17. *https://doi.org/10.1186/1471-2229-11-17*

Engqvist M, Yatusevich R, Müller C, Flügge UI (2018). HAG2/MYB76 and HAG3/MYB29 exert a specific and coordinated control on the regulation of aliphatic glucosinolate biosynthesis in *Arabidopsis thaliana*. New Phytology 177:627-642. *https://doi.org/10.1111/j.1469-8137.2007.02295.x*

Eujayl I, Sledge MK., Wang L, May GD, Chekhovskiy K., Zwonitzer JC (2014). Medicago truncatula EST-SSRs reveal cross-species genetic markers for *Medicago* spp. Theoretical and Applied Genetics 108:414-422. *https://doi.org/10.1007/s00122-003-1450-6*

Feng C, Andreasson E, Maslak A, Mock HP, Mattsson O (2014). Arabidopsis MYB68 in development and responses to environmental cues. Plant Science 167:1099-1107. *https://doi.org/10.1016/j.plantsci.2004.06.014*

Fornalé S, Sonbol FM, Maes T, Capellades M, Puigdomènech P (2016). Down-regulation of the maize and *Arabidopsis thaliana* caffeic acid O-methyl-transferase genes by two new maize R2R3-MYB transcription factors. Plant Molecular Biology 62:809-823. *https://doi.org/10.1007/s11103-006-9058-2*

Fu LG, Jin JM (2012). Red List of Endangered Plants in China. Vol. 1. Science Press.

Fu LG, Li N, Mill RR (2019). Taxaceae. In: Wu ZY, Raven PH (Eds). Flora of China. Beijing: Science Press, pp 89-98. *https://doi.org/10.3389/fgene.2019.01014*

Fu LM, Niu BF, Zhu ZW, Wu ST, Li WZ (2012). CD-HIT: accelerated for clustering the next-generation sequencing data. Bioinformatics 28:3150-3152. *https://doi.org/10.1093/bioinformatics/bts565*

Gao CH, Ren XD, Mason AS, Li J, Wang W, Xiao ML (2013). Revisiting an important component of plant genomes: microsatellites. Functional Plant Biology 40:645-661. *https://doi.org/10.1071/12325*

Garvin DF, McKenzie N, Vogel JP, Mockler TC, Blankenheim ZJ, Wright J, … Shape JW (2010). An SSR-based genetic linkage map of the model grass *Brachypodium distachyon*. Genome 53:1-13. *https://doi.org/10.1139/g09-079*

Gupta PK, Varshney RK (2020). The development and use of microsatellite markers for genetic analysis and plant breeding with emphasis on bread wheat. Euphytica 113:163-185. *https://doi.org/10.1023/A:1003910819967*

Hancock JM (2015). The contribution of slippage-like processes to genome evolution. Journal of Molecular Evolution 41:1038-1047. https://doi.org/10.1007/BF00173185

Haseman JK. Elston RC (2012). The investigation of linkage between a quantitative trait and a marker locus. Behavioral Genetics 2:3-19. https://doi.org/10.1007/BF01066731

He L, Du C, Covaleda L, Xu Z, Robinson AF, Yu JZ, Kohel RJ, Zhang HB (2014). Cloning, characterization, and evolution of the NBSLRR-encoding resistance gene analogue family in polyploidy cotton (*Gossypium hirsutum* L.). Molecular Plant Microbe Interaction 17:1234-1241. https://doi.org/10.1094/MPMI.2004.17.11.1234

Jia XP, Shi YS, Song YC (2017). Development of EST–SSR in foxtail millet (*Setaria italica*) [J]. Genetic Resource and Crop Evolution 54:233-236.

Jiang C, Gu X, Peterson T (2014). Identification of conserved gene structures and carboxy-terminal motifs in the Myb gene family of *Arabidopsis* and *Oryza sativa* L. ssp. *indica*. Genome Biology 5:R46. https://doi.org/10.1186/gb-2004-5-7-r46

Jin H, Cominelli E, Bailey P, Parr A, Mehrtens F (2015). Transcriptional repression by AtMYB4 controls production of UV–protecting sunscreens in Arabidopsis. EMBO Journal 19:6150-6161.

Kaneko M, Inukai Y, Ueguchi-Tanaka M, Itoh H, Izawa T (2014). Loss-of-function mutations of the rice GAMYB gene impair alpha-amylase expression in aleurone and flower development. Plant Cell 16:33-44.

Kantety RV, La Rota M, Matthews DE, Sorrells ME (2012). Data mining for simple sequence repeats in expressed sequence tags from barley, maize, rice, sorghum and wheat. Plant Molecular Biology 48:501-510.

Keller T, Abbott J, Moritz T, Doerner P (2016). Arabidopsis REGULATOR OF AXILLARY MERISTEMS1 controls a leaf axil stem cell niche and modulates vegetative development. Plant Cell 18:598-611.

Koltai H, Bird DM (2011). Epistatic repression of PHANTASTICA and class 1 KNOTTED genes is uncoupled in tomato. The Plant Journal 22:455-459.

Latha MA, Venkateswara Rao K, Dashavantha Reddy V (2015). Production of transgenic plants resistant to leaf blast disease in finger millet (*Eleusine coracana* (L.) Gaertn.). Plant Science 169:657-667.

Laurent, P, Elduque C, Hayes H, Saunier K., Eggen A, Levéziel H (2012). Assignment of 60 human ESTs in cattle Mammal. Genome 11:748-754.

Legay S, Sivadon P, Blervacq AS, Pavy N, Baghdady A (2012). EgMYB1, an R2R3 MYB transcription factor from eucalyptus negatively regulates secondary cell wall formation in Arabidopsis and poplar. New Phytology 188:774-786.

Letunic I, Copley RR, Schmidt S, Ciccarelli FD, Doerks T (2014). SMART 4.0: towards genomic data integration. Nucleic Acids Research 32:D142-D144.

Li L, Darden TA., Weinberg CR, Levine AJ, Pedersen LG (2001). Gene assessment and sample classification for gene expression data using a genetic algorithm/k-nearest neighbor method. Comb. Chem. High Throughput Screening 4:727-739.

Li SF, Milliken ON, Pham H, Seyit R, Napoli R (2019). The Arabidopsis MYB5 transcription factor regulates mucilage synthesis, seed coat development, and trichome morphogenesis. Plant Cell 21:72-89.

Li ZG, Wu MY, Zhao W, Li B, Yang ZP, Hu YM (2003). Detection of 29 types of fusion gene in leukemia by multiplex RT-PCR [in Chinese]. Zhonghua Xue Ye Xue Za Zhi 24:256-258.

Li B, Peng L, Sun X, Huang W, Wang N, He Y, Tang Z (2020). Organ-specific transcriptome sequencing and mining of genes involved in polyphyllin biosynthesis in *Paris polyphylla*. Industrial Crops and Products 156:112775. https://doi.org/10.1016/j.indcrop.2020.112775

Ma QH, Wang C, Zhu HH (2013). TaMYB4 cloned from wheat regulates lignin biosynthesis through negatively controlling the transcripts of both cinnamyl alcohol dehydrogenase and cinnamoyl-CoA reductase genes. Biochemistry 93:1179-1186. https://doi.org/10.1016/j.biochi.2011.04.012

Maia LC, Souza VQ, Kopp MM, Carvalho FIF, Oliveira AC (2019). Tandem repeat distribution of gene transcripts in three plant families. Genetic and Molecular Biology 32:1-12.

Marocco A, Wissenbach M, Becker D, Paz-Ares J, Saedler H (2019). Multiple genes are transcribed in Hordeum vulgare and Zea mays that carry the DNA binding domain of the myb oncoproteins. Molecular Genetics 216:183-187. https://doi.org/10.1007/BF00334354

Martienssen RA, Rabinowicz PD, O'Shaughnessy A, McCombie WR (2014). Sequencing the maize genome. Current Opinion in Plant Biology 7:102-107.

Martins WS, Lucas DCS, Neves KFS, Bertioli DJ (2019). WebSat-A Web Software for microsatellite marker development. Bioinformatics 3(6):28-283. *https://doi.org/10.6026/97320630003282*

Matus JT, Aquea F, Arce-Johnson P (2018). Analysis of the grape MYB R2R3 subgroup reveals expanded wine quality–related clades and conserved gene structure organization across Vitis and Arabidopsis genomes. BMC Plant Biology 8:83-98. *https://doi.org/10.1186/1471-2229-8-83*

Mehrtens F, Kranz H, Bednarek P, Weisshaar B (2015). The Arabidopsis transcription factor MYB12 is a flavonol–specific regulator of phenylpropanoid biosynthesis. Plant Physiology 138:1083-1096. *https://doi.org/10.1104/pp.104.058032*

Mellway RD, Tran LT, Prouse MB, Campbell MM, Constabel CP (2019). The wound-, pathogen-, and ultraviolet B-responsive MYB134 gene encodes an R2R3 MYB transcription factor that regulates proanthocyanidin synthesis in poplar. Plant Physiology 150:924-941. *https://doi.org/10.1104/pp.109.139071*

Meyers BC, Dickerman AW, Michelmore RW, Sivaramakrishnan S, Sobral BW, Young ND (2019). Plant disease resistance genes encode members of an ancient and diverse protein family within the nucleotide-binding superfamily. The Plant Journal 20:317-332. *https://doi.org/10.1046/j.1365-313X.1999.00606.x*

Meyers BC, Kozik A, Griego A, Kuang H, Michelmore RW (2013). Genome-wide analysis of NBS–LRR encoding genes in *Arabidopsis*. Plant Cell 15:809-834. *https://doi.org/10.1105/tpc.009308*

Millar AA, Gubler F (2015). The Arabidopsis GAMYB-like genes, MYB33 and MYB65, are microRNA-regulated genes that redundantly facilitate anther development. Plant Cell 17:705-721. *https://doi.org/10.1105/tpc.104.027920*

Mishra RK, Gangadhar BH, Jae Woong Yu, Doo Hwan Kim, Se Won Park (2011). Development and characterization of EST based SSR markers in Madagascar periwinkle (*Catharanthus roseus*) and their transferability in other medicinal plants. Plant Omics Journal 4(3):154-162.

Morimoto R, Nishioka E, Murai K, Takumi S (2019). Functional conservation of wheat orthologs of maize rough sheath1 and rough sheath2 genes. Plant Molecular Biology 69:273-285. *https://doi.org/10.1007/s11103-008-9422-5*

Müller D, Schmitz G, Theres K (2016). Blind homologous R2R3 Myb genes control the pattern of lateral meristem initiation in Arabidopsis. The Plant Cell 18:586-597. *https://doi.org/10.1105/tpc.105.038745*

Murray F, Kalla R, Jacobsen J, Gubler F (2013). A role for HvGAMYB in anther development. The Plant Journal 33:481-491. *https://doi.org/10.1046/j.1365-313X.2003.01641.x*

Murray MG, Thompson WF (2011). Rapid isolation of high molecular weight plant DNA. Nucleic Acids Research 8:4321-4326. *https://doi.org/10.1093/nar/8.19.4321*

Nesi N, Jond C, Debeaujon I, Caboche M, Lepiniec L (2018). The Arabidopsis TT2 gene encodes an R2R3 MYB domain protein that acts as a key determinant for proanthocyanidin accumulation in developing seed. The Plant Cell 13:2099-2114. *https://doi.org/10.1105/TPC.010098*

Nicholas KB, Nicholas HBJ, Deerfield DWI (2017). Genedoc: analysis and visualization of genetic variation. Embnew News 4:14.

Noda K, Glover BJ, Linstead P, Martin C (2014). Flower colour intensity depends on specialized cell shape controlled by a Myb-related transcription factor. Nature 369:661-664.

Ogata K, Morikawa S, Nakamura H, Hojo H, Yoshimura S (2014). Comparison of the free and DNA–complexed forms of the DNA–binding domain from c–Myb. Nature Structural Biology 2:309-320.

Okuyama Y, Kanzaki H, Abe A, Yoshida K, Tamiru M, Saitoh H, … Terauchi R (2011). A multifaceted genomics approach allows the isolation of the rice Pia-blast resistance gene consisting of two adjacent NBS-LRR protein genes. The Plant Journal 66:467-479.

Palmer LE, Rabinowicz PD, O'Shaughnessy AL, Balija VS, Nascimento LU (2013). Maize genome sequencing by methylation filtration. Science 302:2115-2117.

Payne CT, Zhang F, Lloyd AM (2012). GL3 encodes a bHLH protein that regulates trichome development in *Arabidopsis* through interaction with GL1 and TTG1. Genetics 156:1349-1362.

Paz-Ares J, Ghosal D, Wienand U, Peterson PA, Saedler H (2017). The regulatory c1 locus of maize encodes a protein with homology to myb proto–oncogene products and with structural similarities to transcriptional activators. EMBO Journal 6:3553-3558.

Rabinowicz PD, Braun EL, Wolfe AD, Bowen B, Grotewold E (2019). Maize R2R3 Myb genes: sequence analysis reveals amplification in the higher plants. Genetics 153:427-444.

Perez-Rodriguez M, Jaffe FW, Butelli E, Glover BJ, Martin C (2015). Development of three different cell types is associated with the activity of a specific MYB transcription factor in the ventral petal of *Antirrhinum majus* flowers. Development 132:359-370.

Pilu R, Piazza P, Petroni K, Ronchi A, Martin C (2013). pl-bol3, a complex allele of the anthocyanin regulatory pl1 locus that arose in a naturally occurring maize population. The Plant Journal 36:510-521.

Pletscher-Frankild S, Pallejà A. Tsafou K, Binder JX, Jensen LJ (2015). DISEASES: Text mining and data integration of disease–gene associations. Methods 74:83-89.

Pmaniatis T, Sambrook J, Fritsch EF (2019). Molecular cloning: a laboratory manual, 2nd edn. Cold Spring Harbor Laboratory, New York.

Poncet V, Rondeau M, Tranchant C, Cayrel A, Hamon S, de Kochko A, Hamon P (2016). SSR mining in coffee tree EST databases: potential use of EST-SSRs as markers for the *Coffea* genus. Molecular Genetics and Genomics 276:436-449.

Preston J, Wheeler J, Heazlewood J, Li SF, Parish RW (2014). AtMYB32 is required for normal pollen development in *Arabidopsis thaliana*. The Plant Journal 40:979-995. *https://doi.org/10.1111/j.1365-313X.2004.02280.x*

Province MA, Shannon WD, Rao DC (2001). Classification methods for confronting heterogeneity. Advance Genetics 42:273-286. *https://doi.org/10.1016/S0065-2660(01)42028-1*

PS, Ware D, Fulton RS, Stein JC, Wei F (2019). The B73 maize genome: complexity, diversity, and dynamics. Science 326:1112-1115. *https://doi.org/10.1126/science.1178534*

Qi J, Zheng N, Zhang B, Sun P, Hu S, Xu W, Li X (2013). Mining genes involved in the stratification of *Paris polyphylla* seeds using high-throughput embryo Transcriptome sequencing. BMC Genomics 14(1):1-14. *https://doi.org/10.1186/1471-2164-14-358*

Reddy BLIN, Lakshmi NM, Sivaramakrishnan S (2012). Identification and characterization of EST–SSRs in finger millet (*Eleusine coracana* (L.) Gaertn.). Journal of Crop Science and Biotechnology 15(1):9-16. *https://doi.org/10.1007/s12892-011-0064-9*

Rhoades MW, Reinhart BJ, Lim LP, Burge CB, Bartel B (2012). Prediction of plant microRNA targets. Cell 110:513-520. *https://doi.org/10.1016/S0092-8674(02)00863-2*

Saitoh K, Onishi K, Mikami I, Thidar K, Sano Y (2014). Allelic diversification at the C (OsC1) locus of wild and cultivated rice: nucleotide changes associated with phenotypes. Genetics 168:997-1007. *https://doi.org/10.1534/genetics.103.018390*

Schmitz G, Tillmann E, Carriero F, Fiore C, Cellini F (2012). The tomato Blind gene encodes a MYB transcription factor that controls the formation of lateral meristems. Proceedings of the National Academy of Science 99:1064-1069. *https://doi.org/10.1073/pnas.022516199*

Senthilvel S, Jayashree B, Mahalakshmi V, Sathish Kumar P, Nakka S, Nepolean T, Hash CT (2018). Development and mapping of simple sequence repeat markers for pearl millet from data mining of expressed sequence tags. BMC Plant Biology 8:119. *https://doi.org/10.1186/1471-2229-8-119*

Shannon WD, Province MA, Rao DC (2001). Tree-based recursive partitioning methods for subdividing sibpairs into relatively more homogeneous subgroups. Genetics Epidemiology 20:293-306. *https://doi.org/10.1002/gepi.1*

Sonbol FM, Fornale S, Capellades M, Encina A, Tourin S (2014). The maize ZmMYB42 represses the phenylpropanoid pathway and affects the cell wall structure, composition and degradability in *Arabidopsis thaliana*. Plant Molecular. Biology 70:283-296. *https://doi.org/10.1007/s11103-009-9473-2*

Sønderby IE, Hansen BG, Bjarnholt N, Ticconi C, Halkier BA (2017). A systems biology approach identifies a R2R3 MYB gene subgroup with distinct and overlapping functions in regulation of aliphatic glucosinolates. PLoS One 19:e1322. *https://doi.org/10.1371/journal.pone.0001322*

Squirrell J, Hollingsworth PM, Woodhead M, Russell J, Lowe AJ, Gibby M, Powell W (2013). How much effort is required to isolate nuclear microsatellites from plants? Molecular Ecology 12:1339-1348. *https://doi.org/10.1046/j.1365-294X.2003.01825.x*

Stelzer G, Rosen N, Plaschkes I, Zimmerman S, Twik M, Fishilevich S, Lancet D (2016). The GeneCards suite: from gene data mining to disease genome sequence analyses. Current Protocol in Bioinformatics 54(1):1-30. *https://doi.org/10.1002/cpbi.5*

Stracke R, Ishihara H, Huep G, Barsch A, Mehrtens F (2017). Differential regulation of closely related R2R3–MYB transcription factors controls flavonol accumulation in different parts of the *Arabidopsis thaliana* seedling. The Plant Journal 50:660-677. *https://doi.org/10.1111/j.1365-313X.2007.03078.x*

Szabo A, Boucher K, Jones D, Tsodikov AD, Klebanov LB, Yakovlev AY (2003). Multivariate exploratory tools for microarray data analysis. Biostatistics 4:555-567. *https://doi.org/10.1093/biostatistics/4.4.555*

Thiel T, Michalek W, Varshney RK, Graner A (2013). Exploiting EST database for the development and characterization of gen-derived SSR-markers in barley (*Hordeum vulgare* L.). Theoretical and Applied Genetics 106:411-422. *https://doi.org/10.1007/s00122-002-1031-0*

Thompson JD, Gibson TJ, Plewniak F, Jeanmougin F, Higgins DG (2017). The CLUSTAL_X windows interface: flexible strategies for multiple sequence alignment aided by quality analysis tools. Nucleic Acids Research 25:4876-4882. *https://doi.org/10.1093/nar/25.24.4876*

Tohge T, Nishiyama Y, Hirai MY, Yano M, Nakajima J (2015a). Functional genomics by integrated analysis of metabolome and transcriptome of Arabidopsis plants over-expressing an MYB transcription factor. The Plant Journal 42:218-235. *https://doi.org/10.1111/j.1365-313X.2005.02371.x*

Tsamardinos I, Aliferis CF (2003). Towards principled feature selection: relevance, filters and wrappers. In: Ninth International Workshop on Artificial Intelligence and Statistics, Key West, FL.

Tsiantis M, Schneeberger R, Golz JF, Freeling M, Langdale JA (2020). The maize rough sheath2 gene and leaf development programs in monocot and dicot plants. Science 284:154-156. *https://doi.org/10.1126/science.284.5411.154*

Tsuji H, Aya K, Ueguchi-Tanaka M, Shimada Y, Nakazono M (2016). GAMYB controls different sets of genes and is differentially regulated by microRNA in aleurone cells and anthers. The Plant Journal 47:427-444. *https://doi.org/10.1111/j.1365-313X.2006.02795.x*

Van Aalten DM, Grotewold E, Joshua-Tor L (2018). Essential dynamics from NMR clusters: dynamic properties of the Myb DNA-binding domain and a hinge-bending enhancing variant. Methods 14:318-328. *https://doi.org/10.1006/meth.1998.0587*

Varshney RK, Thiel T, Stein N, Langridge P, Graner A (2012). *In silico* analysis on frequency and distribution of microsatellites in ESTs of some cereal species. Cellular and Molecular Biology Letters 7:537-546.

Victoria FC, da Maia Luciano C, de Oliveira Antonio C (2011). *In silico* comparative analysis of SSR markers in plants. BMC Plant Biology 11:15–30. *https://doi.org/10.1186/1471-2229-11-15*

Vizir I, Twell D (2015). Male germ line development in *Arabidopsis*: duo pollen mutants reveal gametophytic regulators of generative cell cycle progression. Plant Physiology 137:297-307. *https://doi.org/10.1104/pp.104.053165*

Wagner AH, Coffman AC, Ainscough BJ, Spies NC, Skidmore ZL, Campbell KM, Griffith OL (2016). DGIdb 2.0: mining clinically relevant drug–gene interactions. Nucleic Acids Research 44(D1): D1036-D 1044. *https://doi.org/10.1093/nar/gkv1165*

Waites R, Selvadurai HR, Oliver IR, Hudson A (2018). The PHANTASTICA gene encodes a MYB transcription factor involved in growth and dorsoventrality of lateral organs in Antirrhinum. Cell 93:779-789. *https://doi.org/10.1016/S0092-8674(00)81439-7*

Walker AR, Davison PA, Bolognesi-Winfield AC, James CM, Srinivasan N (2020). The Transparent Testa Glabra1 locus, which regulates trichome differentiation and anthocyanin biosynthesis in Arabidopsis, encodes aWD40 repeat protein. The Plant Cell 11:1337-1350. *https://doi.org/10.1105/tpc.11.7.1337*

Watanabe N, Kobayashi H, Ichiji O, Yoshida MA, Kikuta A, Komada Y, Sekine I, Ishida Y, Horiukoshi Y, Tsunematsu Y (2003). Cryptic insertion and translocation or nondividing leukemic cells disclosed by FISH analysis in infant acute leukemia with discrepant molecular and cytogenetic findings. Leukemia 17:876-882. *https://doi.org/10.1038/sj.leu.2402900*

Weising K, Nybom H, Wolff K, Kahl G (2015). Application of DNA fingerprinting in plant sciences. DNA Fingerprinting in plants-principles, methods, and applications. CRC Press, Boca Raton, pp 235-276. *https://doi.org/10.1093/aob/mcj057*

Wilkins O, Nahal H, Foong J, Provart NJ, Campbell MM (2019). Expansion and diversification of the Populus R2R3-MYB family of transcription factors. Plant Physiology 149:981-993. *https://doi.org/10.1104/pp.108.132795*

Williams CE, Grotewold E (2017). Differences between plant and animal Myb domains are fundamental for DNA binding activity, and chimeric Myb domains have novel DNA binding specificities. Journal of Biological Chemistry 272:563-571. *https://doi.org/10.1074/jbc.272.1.563*

Woodger FJ, Gubler F, Pogson BJ, Jacobsen JV (2013). A Mak-like kinase is a repressor of GAMYB in barley aleurone. The Plant Journal 33:707-717. *https://doi.org/10.1046/j.1365-313X.2003.01663.x*

Xiaoyuan Y, Kun H, Meihua L, Jigang L (2016). The MYB transcription factor superfamily of Arabidopsis: expression analysis and phylogenetic comparison with the rice MYB family. Plant Molecular Biology 60:107-124. *https://doi.org/10.1007/s11103-005-2910-y*

Xu R, Wang Z, Su Y, Wang T (2020). Characterization and development of microsatellite markers in *Pseudotaxus chienii* (Taxaceae) based on transcriptome sequencing. Frontier in Genetics 11:1249. *https://doi.org/10.3389/fgene.2020.574304*

Yu JK, Dake TM, Singh S, Benscher D, Li W, Gill B, Sorrells ME (2014). Development and mapping of EST-derived simple sequence repeat markers for hexaploid wheat. Genome 47:805-818.

Zane L, Bargelloni L, Patarnello T (2012). Strategies for microsatellite isolation: a review. Molecular Ecology 11:1-16. *https://doi.org/10.1046/j.0962-1083.2001.01418.x*

Zeng S, Xiao G, Guo J, Fei Z, Xu Y, Bruce A Roe, Wang Y (2010). Development of a EST dataset and characterization of EST-SSRs in a traditional Chinese medicinal plant, *Epimedium sagittatum* (Sieb. Et Zucc.) maxim. BMC Genomics 11:94-105. *https://doi.org/10.1186/1471-2164-11-94*

Zhang F, Gonzalez A, Zhao M, Payne CT, Lloyd A (2013). A network of redundant bHLH proteins functions in all TTG1-dependent pathways of Arabidopsis. Development 130:4859-4869. *https://doi.org/10.1242/dev.00681*

Zhang F, Peterson T (2015). Comparisons of maize pericarp color1 alleles reveal paralogous gene recombination and an organ-specific enhancer region. Plant Cell 17:903-914. *https://doi.org/10.1105/tpc.104.029660*

Zhang L, Chia JM, Kumari S, Stein JC, Liu Z (2019). A genome-wide characterization of microRNA genes in maize. PLoS Genetics 5:e1000716. *https://doi.org/10.1371/journal.pgen.1000716*

Zhang L, Zhao G, Jia J, Liu X, Kong X (2012). Molecular characterization of 60 isolated wheat MYB genes and analysis of their expression during abiotic stress. Journal of Experimental Botany 63:203-214. *https://doi.org/10.1093/jxb/err264*

Zimmermann IM, Heim MA, Weisshaar B, Uhrig JF (2014). Comprehensive identification of Arabidopsis thaliana MYB transcription factors interacting with R/B-like BHLH proteins. The Plant Journal 40:22-34. *https://doi.org/10.1111/j.1365-313X.2004.02183.x*