



Understanding saffron biology using omics- and bioinformatics tools: stepping towards a better *Crocus* phenome

Amjad M. Husaini¹ · Syed Anam ul Haq¹ · Alberto José López Jiménez²

Received: 16 September 2021 / Accepted: 6 December 2021 / Published online: 2 February 2022
© The Author(s), under exclusive licence to Springer Nature B.V. 2021

Abstract

Saffron is a unique plant in many aspects, and its cellular processes are regulated at multiple levels. The genetic makeup in the form of eight chromosome triplets ($2n = 3x = 24$) with a haploid genetic content (genome size) of 3.45 Gbp is decoded into different types of RNA by transcription. The RNA then translates into peptides and functional proteins, sometimes involving post-translational modifications too. The interactions of the genome, transcriptome, proteome and other regulatory molecules ultimately result in the complex set of primary and secondary metabolites of saffron metabolome. These complex interactions manifest in the form of a set of traits ‘phenome’ peculiar to saffron. The phenome responds to the environmental changes occurring in and around saffron and modify its response in respect of growth, development, disease response, stigma quality, apocarotenoid biosynthesis, and other processes. Understanding these complex relations between different yet interconnected biological activities is quite challenging in saffron where classical genetics has a very limited role owing to its sterility, and the absence of a whole-genome sequence. Omics-based technologies are immensely helpful in overcoming these limitations and developing a better understanding of saffron biology. In addition to creating a comprehensive picture of the molecular mechanisms involved in apocarotenoid synthesis, stigma biogenesis, corm activity, and flower development, omics-technologies will ultimately lead to the engineering of saffron plants with improved phenome.

Keywords *Crocus sativus* L. · Saffron · Genomics · Transcriptomics · Proteomics · Metabolomics · Bioinformatics

Introduction

Saffron (*Crocus sativus* L.) is a sterile triploid plant characterized by its red stigmas which, when desiccated, constitute the spice known as saffron. The word “saffron” is derived from “zafran”, the Arabic word that translates to “yellow”. The *Crocus* genus comprises over 200 recognized species of perennial geophytes. It has been proposed that *C. sativus* L. is an autotriploid hybrid derived from *C. cartwrightianus*. *C. sativus* L. belongs to the Iridaceae (Liliales, Monocots)

whose genomes are relatively large. From the point of view of agricultural production, *Crocus sativus* L. is propagated vegetatively using corms, and is prevalent throughout the tropical and subtropical regions of the northern hemisphere [1]. The major saffron growing regions of the world include Iran, Azerbaijan, Spain, Italy, India (Kashmir), Greece, and Turkey. The total world saffron production was estimated at 378.33 tons in 2016 [2], of which about 90% is produced in Iran whereas the rest of the production is located in India (Kashmir), Greece, Afghanistan, Spain and Italy [2].

Crocus sativus L. genome comprises eight chromosome triplets ($2n = 3x = 24$), with a genome size of $1C = 3.45$ Gbp [3]. Several reports have highlighted saffron countless medicinal properties like anticancer, antimutagenic and antioxidant effects [4]. Saffron bioactive compounds have immense therapeutic properties useful for coronary artery diseases, neurodegenerative disorders, bronchitis, asthma, diabetes, fever, and colds. It has the potential to help tackle problems associated with severe acute respiratory syndrome (COVID-19) patients and post-covid-19 problems [4]. It can help manage stress and anxiety during isolation, quarantine

✉ Amjad M. Husaini
amjadhusaini@skuastkashmir.ac.in

¹ Genome Engineering and Societal Biotechnology Lab, Division of Plant Biotechnology, Sher-e-Kashmir University of Agricultural Sciences and Technology of Kashmir, Shalimar Campus, Srinagar, Jammu and Kashmir 190025, India

² Departamento de Ciencia y Tecnología Agroforestal y Genética, Escuela Técnica Superior de Ingenieros Agrónomos y de Montes, Universidad de Castilla-La Mancha, Albacete, Spain

and lockdowns. Its efficacy in managing depression is comparable to drugs like imipramine, fluoxetine, and citalopram. Owing to these properties and the glamour associated with it, saffron is one of the costliest spices in the world.

Saffron is propagated through corms [5], and does not produce fertilisable gametes [6] and is self-incompatible [7]. This makes all modern saffron plants almost genetically identical. This is a bottleneck for the genetic improvement of this highly valued crop. The omics-based studies can be a benchmark for its genetic improvement [8]. These studies in saffron have broadly focused on the below-mentioned research areas.

Flower development and stigma apocarotenoid content

The most valued metabolites in *Crocus sativus* are synthesized in stigma tissue in a developmental stage-specific manner. Different stages are defined according to stigma development based on the length of the stigma, its pigmentation, and apocarotenoid content (generally yellow, orange and red in the preanthesis stigma). Almost a decade ago, we highlighted the importance of the saffron stigma transcriptome characterization for understanding the molecular basis of its flavour and colour biogenesis, the gynoeceum developmental biology, and its genomic organization [9, 10]. We expected functional genomics of *Crocus sativus* to play a vital role in finding candidate genes for producing stigma pigments and flavouring compounds. This would enable overexpression studies on saffron for enhancing the production of these pigments and flavouring compounds, and improve the quality of saffron.

Besides whole-genome sequencing, expressed sequence tags (ESTs) are a vital source for analyzing gene expression in specific organs, growth stages, developmental processes, and stress response in crops [10]. The first important database of ESTs for stigma biogenesis and apocarotenoid pathway contained 6768 ESTs [11]. The most relevant contigs included those encoding non-heme- β -carotene-hydroxylase, putative glycosyltransferase, putative isoprenoid GTases, Myb-like protein, Myb305, and Cytochrome P450 [9]. Analysis of saffron stigma EST collections at different developmental stages revealed that CsCCD2 (carotenoid cleavage dioxygenase) ESTs are predominant in the early stages [12].

Transcriptome analyses in saffron and related species (including leaves, stamens, corm, tepals, and stigmas) have uncovered a large number of transcription factor-coding genes [13–15]. The main transcripts encoding TFs upregulated in stigma are those involved in secondary metabolite biosynthesis. Among them, transcripts encoding MYB, MYB related, WRKY, C2C2-YABBY and bHLH transcription factors have been shown to be differentially expressed

[15]. Besides, a total of 1075 transcripts revealed a tissue-specific expression, out of which 342 were expressed in stamen, 304 in leaf, 161 in tepal, 144 in stigma and 124 in corm.

Using deep transcriptome analysis, a novel dioxygenase carotenoid cleavage dioxygenase (CCD2), which catalyzes the first step of crocin biosynthesis from carotenoid zeaxanthin, has been identified [12]. Other studies have led to the characterization of the regulation of crocetin and crocins biosynthesis from phytoene [16]. A recent study has identified a new glycosyltransferase, UGT91P3, responsible for the last glycosylation step in the biosynthesis of crocins [17].

Genes encoding enzymes for volatile biosynthesis have been identified using in silico screening of the stigma cDNA database described previously [11, 18]. Comparison of the apocarotenoid content and expression profiles show that 1 deoxyxylulose 5 phosphate synthase (DXS) plays a vital role in apocarotenoid accumulation. DXS is expressed at all the developmental stages of *C. sativus* stigma, while 3 hydroxy 3 methylglutaryl CoA reductase (HMGR) is expressed at low levels only. Additionally, two putative terpene synthases (TS1 and TS2) showed differential expression, with TS2 having an important role in the biosynthesis of apocarotenoids. The expression of two carotenoid biosynthesis genes, CsPSY (phytoene synthase) and CsPDS (phytoene desaturase), increases in the red stage. In another study, it was observed that during the transition from yellow to red stigmas, accumulation of zeaxanthin was accompanied by enhanced expression of phytoene synthase, phytoene desaturase and lycopene cyclase [19]. Besides, a massive accumulation of carotene hydroxylase and zeaxanthin cleavage dioxygenase transcripts was described (Fig. 1).

A systematic comparative analysis of crocin data and transcriptomes from *C. sativus*, *C. ancycensis* and *C. cartwrightianus*, has led to the identification of putative transcription factors affecting apocarotenoid accumulation during stigma development in saffron [20]. Expression levels of DXS-CLA1, ZDS, Z-ISO, PDS, CrtISO, BCH-2, LYC-B, CCD2, and UGT74AD2 had a positive correlation with apocarotenoid levels in the three species. Besides, in stigma, eleven TFs belonging to the bHLH, C2H2, ARF, HB, CBF/DREB1, NF-YC and ALFIN families showed correlation between expression and apocarotenoid levels in the three species. A similar study [21] compared the transcriptomes of cultivated *C. sativus* and wild *C. cartwrightianus*. This study found seven genes related to apocarotenoid biosynthesis, which showed differential expression between the samples. These genes are orthologues of carotenoid isomerase (CsTc091265), lycopene beta-cyclase (CsTc018497), zeaxanthin epoxidase (CsTc006236), UDP-glycosyltransferase (CsTc020060), phytoene synthase (CsTc009491), nine-cis-epoxy carotenoid dioxygenase (CsTc035409), and carotene beta-hydroxylase (CsTc000418). These findings provide important information for the saffron improvement

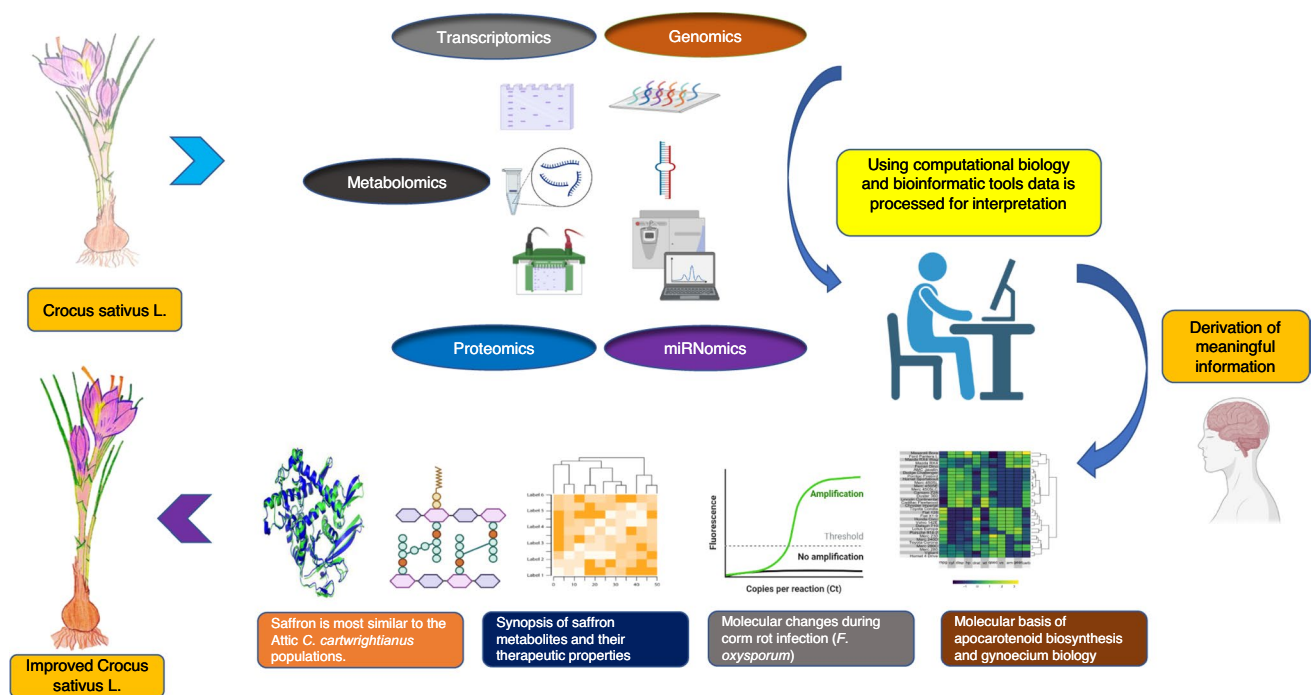


Fig. 1 Major characteristics studied using omics-based approaches for understanding saffron biology

program. The orthologue of gene UDP-glucosyltransferase (CsTc020060) was down-regulated in all individual saffron plants while it is up-regulated in all the *C. cartwrightianus* plants [21]. UDP-glucosyltransferase, being involved in the conversion of crocetin to crocin, could be a cause behind the difference in metabolite accumulation between *Crocus* species. Since triploidy and sterility help safeguard the favourable allele composition (regarding aroma and colour) from being segregated by recombination, modulation of gene expression using genome modification and advanced genetic engineering approaches can be a smart strategy to increase saffron apocarotenoid content in stigma, improving saffron quality and enhancing its economic value.

Understanding saffron flower development is vital for improving its productivity and quality. The combination of class A genes (including APETALA1; CsAP1 and APETALA2; CsAP2), class B genes (including APETALA3; CsAP3 and PISTILLATA; CsPI) and class C genes (including AGAMOUS; CsAG), determines the identity of the organs developing in a whorl. An important gene in the stigma development of saffron is the C-class floral homeotic gene AGAMOUS (CsAG) [22]. Its expression began in the yellow stage of stigma, showed 16 folds increase as stigma turned from yellow to the orange stage and continued to increase up to the scarlet stage [22]. Similarly, the expression of the transcript UGT85U1 increases from yellow stage to red stage and anthesis. However, CsNCED, a regulatory gene encoding the enzyme involved in ABA biosynthesis

shows lower expression in all the developmental stages [23]. Relative transcript changes of CsAP3 and NAC-like protein (CsNAP) genes have also been studied during different stages of flower development [24, 25]. However, no direct correlation in the expression of these genes could be detected. CsAP3 expression was maximum during the late pre-anthesis of stigma development, while CsNAP expression increased abruptly at the scarlet stage of stigma. The study concluded that some factor(s) could regulate CsNAP expression, while CsAP3 gene could in turn regulate the factor(s). The promoter of CsAP3 gene consists of three CARG regions, which play a pivotal role in the expression of AP3 gene, of which CARG1 is the binding site for activator proteins, thus regulating floral growth. Given this, Wafai et al. [25] conducted a study to understand the interaction between nuclear factors with B class gene CsAP3 through its CARG1 promoter region. Nuclear proteins were isolated, and a CARG1 sequence was synthesized artificially. Using Electrophoretic Mobility Shift Assay (EMSA), the binding interaction of CARG1 region with pure nuclear protein was studied, and the complex was used for protein identification using LCMS. CsNAP was identified as a conspicuous homeotic protein interacting with CARG1 region of AP3 promoter. Understanding the pathway and deciphering the complete mechanism of floral organ differentiation can pave the way for prolonged flowering of saffron by artificially manipulating the key genes. It will provide farmers ample time to collect the flowers and regulate flowering time/duration so

that flower damage caused due to early frost in November can be avoided.

In a step forward to better understand the flowering mechanism, two sets of full-length transcriptomes of flowering and non-flowering saffron crocus have been generated using NGS and SMRT sequencing [26]. Recently, morphological, physiological and transcriptome analyses of apical bud samples of *C. sativus* were performed during the floral transition process, and a hypothetical model for the regulatory networks of the saffron flowering transition was proposed [27].

Proteomics is central to the understanding of saffron biology. However, not much work has been reported in saffron proteomics, unfortunately. Not many data sets are available in the PRIDE PRoteomics IDentifications (PRIDE) Archive database [28]. The first dedicated protein database for saffron stigma (*Crocus sativus* L, taxonomy-id: 82528) samples at different developmental stages have been created only in the recent past [29]. The MS proteomics data can be accessed from the ProteomeXchange Consortium via the PRIDE partner repository with the data set identifier PXD009014 (<https://www.ebi.ac.uk/pride/archive/projects/PXD009014>). In another recent study, protein profiling of flowering and non-flowering saffron buds subjected to cold stress was done using isobaric tags for relative or absolute quantitation (iTRAQ). Out of 5624 proteins identified in the study, 201 were differentially abundant protein species (DAPs) between these two groups. Upregulated DAPs play an important role in sucrose metabolism, lipid transport, glutathione metabolism, and gene silencing by RNA. Downregulated DAPs are involved in starch biosynthesis and oxidative stress response. Three new flower-related proteins, CsFLK, CseIF4a, and CsHUA1 were identified too [30].

A search in the GenBank protein database for saffron leads to just over 530 entries, these entries correspond to *C. sativus* (268), *C. cartwrightianus* (258), and *C. ancyrensis* (4) (<http://www.ncbi.nlm.nih.gov/>). Despite several tools available for predicting and visualising secondary and tertiary structures of proteins, there is no detailed analysis in saffron. A search on saffron crocus query in the UniProt Knowledgebase (UniProtKB) returns only 426 entries, out of which 420 are in Unreviewed (TrEMBL). Only six have been manually reviewed in Swiss-Prot, and include Crocetin glucosyltransferase 2, Crocetin glucosyltransferase 3, Profilin, Zeaxanthin 7,8(7',8')-cleavage dioxygenase, Carotenoid 9,10(9',10')-cleavage dioxygenase, and Pollen allergen Cro s 1. We could find only one 3-D X-ray diffraction-based crystal structure of saffron protein in the protein data bank (PDB) viz. Cysteine Protease (at 1.3 Å Resolution) and is available at <http://www.rcsb.org/pdb/explore/explore.do?structureId=3U8E>.

As already highlighted, unlike rice, maize, wheat, or tomato, there are limited saffron-specific genomic resources available to explore its peculiar biology. There is a need to

explore and utilize most modern technologies that can generate maximum useful information. Activity-based protein profiling (ABPP) is one such novel techniques of chemical proteomics that has recently revolutionized proteomics. Besides its use in drug selectivity and diagnostics, its application in plant science has grown in the last few years [31]. ABPP uses small molecules as probes for labelling enzymes when these are in an active state. In saffron, the first report on ABPP demonstrated the multiplexing of probes and generated useful information about the active proteases involved in the different developmental stages of stigma [29]. This approach has successfully identified and quantified sixty-seven differentially active glycosidases during the stigma development, implying that glycosidase activity is vital for stigma maturation. These results suggested potential candidate glycosidases involved in the conversion of picrocrocin into safranal.

GOLM and the MASSBANK are popular databases for metabolomic profiling, providing information of reference mass spectra from biologically active metabolites. Other databases like Kyoto Encyclopedia of Genes and Genomes (KEGG), Reactome, MetaCyc and GO-ontology are important resources for curated information regarding biochemical pathways and molecular functions wherein these metabolites perform specific roles. Studying the flavonoid glucosylation and carotenoid biosynthesis enzyme metabolomics [18, 32] is vital for understanding the dynamics of these pathways in saffron. Metabolic analysis of stigma at yellow stage has shown low levels of crocetin, crocins, picrocrocin, and some unidentified compounds with maximum wavelengths around 250 nm. Picrocrocin and crocins have been detected early in the orange stage, increasing rapidly in the red stage. The glycosylated products of crocetin reach maximum levels in the red stage [33]. Picrocrocin level rises in the orange stage and achieves the maximum level at anthesis [34].

Besides apocarotenoids, saffron contains volatile compounds also. More than 160-volatile compounds have been detected using chromatography, spectroscopy and mass spectrometry techniques [35, 36]. In the yellow stigma (stage), the fatty acid derivatives predominate, while in the orange (stage), carotenoid derivatives too are present in addition to the fatty acid derivatives. In the red stigma (stage), the volatiles derived from carotenoids accumulate to high levels, and β -cyclocitral, generated by the cleavage of β -carotene reaches maximum level. Just before anthesis at the scarlet stigma (stage), the volatile propanoic acid, 2-methyl-2,2-dimethyl-1-(2-hydroxy-1-methylethyl) propyl ester accumulate at high levels. However, their levels decrease at anthesis when monoterpenes and carotenoids reach their maximum levels [10]. Among the monoterpenes, linalool is emitted at high levels at anthesis and is responsible for floral odours [37]. In the post-anthesis stage, the fatty acid-derived volatiles become the main volatile compounds.

Diversity of saffron and its characterization

Despite the advancement of sequencing technology and its affordability, there is no whole-genome sequence available for any *Crocus* species. Some classical cytogenetic analyses involving chromosome counting and karyotyping have been done in saffron [38]. Those studies have shown that saffron is a triploid with karyotype $2n = 3x = 24$. It comprises of 8 triplets: two triplets are sub-acrocentric, three triplets are metacentric, two triplets are sub-metacentric and one triplet contains two kinds of chromosomes: chromosome 5(1), metacentric, and chromosomes 5(2,3), sub-acrocentric and smaller [10]. Some efforts have been made to improve our understanding of the genomic organisation of *Crocus* species. These studies are mostly based on Random amplified polymorphic DNA (RAPD) [39], Inter-Retrotransposon Amplified Polymorphism (IRAP) markers [40, 41], Amplified Fragment Length Polymorphism (AFLP) and Simple Sequence Repeats (SSR) [42].

The barcode analysis of the 86 species of genus *Crocus* using *rpoC1*, *matK* and *tmH-psbA* regions has shown the importance of barcoding in the genetic diversity of *Crocus* [43]. RAPD and inter simple sequence repeat (ISSR) marker profiles of 43 isolates of *C. sativus* collected from different geographical areas have been used to determine if this species is monomorphic or polymorphic. The results showed that the clones were identical at the molecular level [44]. Surprisingly, ISSR markers showed no differences between *C. sativus* and *C. cartwrightianus* [45]. In contrast, RAPD markers revealed considerable genetic diversity among 10 elite saffron clones selected in Kashmir [46]. Long terminal repeats (LTRs), a retrotransposon (RTN)-based marker study in Iranian species of *Crocus* showed high diversity within and between species [40]. Using 12 microsatellite markers [47] succeeded in detecting good polymorphism within fifty Iranian individuals of *Crocus sativus*. A reasonable amount of polymorphism was detected in similar studies among Iranian *C. sativus* germplasms [48].

There is ample evidence that epigenetics plays an important role in creating inheritable variation contributing significantly to different traits in several plant species. DNA methylation is the most widely studied epigenetic mark in plants as its genome-wide investigation is easier to accomplish [49]. In a study involving more than a hundred saffron accessions from WSCC (World Saffron and *Crocus* Collection, Spain), very low genetic variability was detected using 12 AFLP (Sensitive Amplified Fragment Length Polymorphism) primer combinations. In contrast, very high epigenetic variability was detected with just 3 Methylation-Sensitive Amplified Fragment Length Polymorphism (MS-AFLP) primer combinations [50]. Five

accessions from the WSCC germplasm having extremely low genetic variability were cultivated for 3 years in the same field. These accessions of different origins maintained different epigenotypes. This suggests that the epigenetic structure of saffron is highly stable [51]. The stability of saffron epigenotype over the years supports the idea that epigenetics may play a vital role in the constancy of saffron phenotype variability.

AFLP analysis using methylation-sensitive restriction enzyme-sequencing (MRE-seq) gave more insight into saffron's epigenome [52]. The study compared the epigenetic profile of 5 phenotypically different, but genetically similar accessions from WSCC germplasm. Differential methylation of regions was detected in some genes encoding transcription factors, shaping the alternative phenotypes. Many SNPs and INDELs were identified, showing thereby that genetic polymorphism exists within the saffron species. Genetic variants were also detected in Gene Ontology (GO) terms, portraying a genetic basis for alternative phenotypes. A heatmap of the 50 highest polymorphic GOs shared between accessions highlighted the presence of two distinct clusters of Indian and Spanish accessions. Twelve GOs showed lower polymorphism in the Spanish accessions than Indian accessions [52].

Phylogenetic analyses of nuclear loci and chloroplast genome, as well as genome-wide DNA polymorphism, indicate that *Crocus sativus* is genetically closer to *C. cartwrightianus* populations. Genome sequencing and Fluorescence in situ hybridisation (FISH) have demonstrated that genomes of two *Crocus cartwrightianus* individuals with slight chromosomal differences had gotten fused, and it could be the parental origin of saffron *Crocus sativus* L. [53]. Another view is that the most likely ancestors of saffron are *C. cartwrightianus* and *C. pallasii* subsp. *Pallasii* (or close relatives) [54].

Saffron growth, development and disease

While there are ample omics-based studies on apocarotenoid biosynthesis pathway, the studies on saffron growth and development are limited. Proteomic analysis has led to the identification of differentially accumulated proteins (in somatic embryos) of *C. sativus*. Thirty-six proteins have been identified, including those involved in protein synthesis, carbohydrate and energy metabolism, defence and stress response, nitrogen metabolism, and secondary metabolism [55]. Metabolomic studies have provided insights into the corm composition of *C. sativus*, too [56]. At the sprouting stage (in corms), sugars like glucose, fructose, and maltose reveal a strong positive correlation with palmitate, turanose, oxalic acid, ethanolamine, linoleic acid, and tetronic acid; and a negative correlation with sitosterol mannoside and

octadecanoic acid. At bud development, fatty-acid biosynthesis significantly relied on carbohydrate metabolism intermediates. Sucrose breakdown reached its maximum to begin the sprouting and bud growth in *C. sativus*.

Climate change and the associated biotic and abiotic stresses are the most daunting challenges to saffron cultivation [8]. Omics-based biological studies of saffron crop shall pave the way for its sustainable production, especially given the problems associated with climate change. MicroRNome of plants, though ubiquitous and small in size, plays an important role in abiotic stress. MicroRNA sequencing, yet ignored in *C. sativus*, can be vital in understanding the regulation of saffron genomic elements. These can also throw light on the regulatory networks underlying the apocarotenoid biosynthesis in *C. sativus*. A study on an EST library from mature *C. sativus* stigmas has helped detect two putative microRNAs, miR414 and miR837-5p, in saffron stigma [57]. Co-expression network analysis has revealed them playing vital roles in metabolic pathways. The predicted targets of the miR414 are: β -carbonic anhydrase 5, Transducin/WD40 repeat-like superfamily protein and three-transposable element genes AT2G13700.1, AT4G06613.1, AT3G29783.1. The predicted targets of miR837-5p are SEC14 cytosolic factor family protein/phosphoglyceride transfer family protein, enhancer of polycomb-like protein, and F-box/RNI-like/FBDlike domains containing protein. In addition, three more miRNAs, csa-miR1, csa-miR2 and csamiR3 have also been predicted by using in silico methods of EST analysis [58]. The predicted targets of these miRNAs are involved in regulating plant growth, senescence, stress responses, disease resistance, mRNA export, protein synthesis and post-translational modifications [58].

In an RNA-seq based transcriptome study, useful information was categorised in the form of small databases for viruses, bacteria, fungi, and plants [59]. It used YeATS suite from the NCBI and Ensembl databases, and showed that the soybean mosaic virus is abundantly expressed in the corm, tepal, leaf, stigma, and stamen tissues [59]. Furthermore, it has been shown that there is a difference in fungal diversity between roots and corms of *C. sativus*. At the flowering stage, the dominant phylum in the rhizosphere is Zygomycota, while in the cormosphere Basidiomycota is dominant. In the cormosphere, Basidiomycota is prevalent at the flowering stage, while Zygomycota is dominant at the dormant stage. However, in the bulk soil, Ascomycota dominates during both stages [60].

Saffron corm rot caused by *Fusarium oxysporum* is a major disease, causing heavy losses in saffron-producing countries [8]. ABPP, a chemical proteomics-based technique, has been upscaled by multiplexing diverse probes (targeting serine hydrolases, α -glycosidases, β -glycosidases and cysteine proteases) to give a broad snapshot of active proteases having a role in corm rot infection [29]. The

suppressed activity of an α -glycosidase upon *F. oxysporum* infection has been detected, which is consistent with the view that *F. oxysporum* suppresses AGLU1 in the apoplast to overcome its antifungal activity [29]. While the activities of putative α -glycosidases (100-kD) and β -glucosidases (50–70 kD) increased upon infection, the activities of serine hydrolases (50, 60 kD) decreased. Additionally, many β -glucosidases (45–60 kD) appeared, while some (65–70 kD) disappeared. In the ABPP based chemical proteomics study, drastic changes were visualised in the activity profile of cysteine proteases, especially papain-like Cys proteases and vacuolar processing enzymes (Table 1).

Saffron adulteration and spice quality

The molecular analysis involving a complete set of metabolites existing in a cell at a particular instant is the backbone of understanding metabolic pathways and is called metabolomics. It is highly significant for plants due to the crucial role of the secondary metabolites in plant survival. These metabolites are extracted from the tissues, separated and analysed in a high-throughput manner to generate metabolic fingerprints. Many tools available in the bioinformatics toolbox help identify and characterise these metabolites. In saffron, metabolite fingerprinting (based on ¹H NMR spectra) and chemometrics have helped in authenticating saffron as Italian or Iranian [61]. These techniques also helped to detect the presence of plant-based adulterants in saffron [62]. ¹H NMR and chemometrics studies have shown that saffron can preserve its valuable characteristics up to 4 years [63].

High-performance thin-layer chromatography (HPTLC) helps study chemical diversity among saffron accessions. In a recent study, 53 saffron accessions from Khorasan Razavi were characterized for chemical diversity using HPTLC. Based on the heat maps generated at different wavelengths, crocin and picrocrocin content was found helpful in categorising saffron [64]. The third important bioactive molecule, safranal, is not among the major volatiles produced in the fresh tissue. Safranal, which gets produced by picrocrocin degradation during the dehydration of the stigma, is the primary aroma component comprising 60–70% essential oil content [65].

Medicinal value and drug development

Saffron bioactive compounds have immense therapeutic properties, including those beneficial against coronary artery diseases, neurodegenerative disorders, bronchitis, asthma, diabetes, fever, colds, and metabolic syndrome. A detailed analysis of its medicinal properties points to its immense untapped potential for easing the distress symptoms of

Table 1 Significant research findings and outcomes of omics-based research studies conducted in saffron and its allies

S. no.	Omics approach used	The gist of the main findings and outcomes	References
1.	Genomics	<p>Whole genome sequencing of <i>Crocus</i> sp. has not been done</p> <p>There are contradictory results on the detection of polymorphisms using marker-based analysis</p> <p>Some studies conclude that saffron is a monomorphic species and whole genome sequencing is needed to discriminate between its isolates</p> <p>Some studies show that molecular markers are quite efficient in detecting polymorphism. Such studies conclude that saffron is not monomorphic and that there is diversity which could be useful for breeding purposes</p> <p>AFLP analysis using methylation-sensitive restriction enzyme-sequencing (MRE-seq) has shown that phenotypically different but genetically similar accessions vary in the methylation pattern of genomic regions encoding transcription factors and may result in alternative phenotypes</p> <p>Epigenetic structure in saffron is highly stable and may play a vital role in the constancy of saffron phenotype variability</p> <p>ISSR primers are reported to be capable of easily distinguishing genuine saffron from fake one</p>	[44, 47, 47, 48, 51, 52, 54, 88–91]
2.	Transcriptomics	<p>De novo transcriptome assemblies have been created from leaves, stamens, corm, tepals, and stigmas of <i>Crocus sativus</i></p> <p>The most valued compounds of <i>C. sativus</i> are synthesised inside stigma in a developmental stage-specific manner</p> <p>During the transition from yellow stage to red stage stigmas there is an accumulation of zeaxanthin accompanied by sharp increase in the expression of phytoene synthase, phytoene desaturase, lycopene β cyclase, β carotene hydroxylase and zeaxanthin cleavage dioxygenase</p> <p>CsCCD2 (carotenoid cleavage dioxygenase) ESTs are prominent in the saffron stigma libraries obtained from early stages of stigma development</p> <p>UDP-glucosyltransferase is vital for conversion of crocetin to crocin, and therefore causes difference in metabolite accumulation between <i>Crocus</i> species</p> <p>1 Deoxyxylulose 5 phosphate synthase (DXS) plays a vital role in apocarotenoid accumulation in stigma</p> <p>There is no direct concordance in the expression of CsAP3 and CsNAP gene expression in saffron</p> <p>Identification, isolation, and biochemical characterisation of uridine diphosphate glycosyltransferase (UGT709G1), which catalyses the HTCC glucosyltransferase reaction to yield picrocrocin, can provide a vital lead for the industrial production of picrocrocin/safranal</p> <p>Differentially expressed full-length transcripts of flowering and non-flowering saffron crocus have been identified and characterised</p> <p>Stigma development in field- and indoor-cultivated saffron is similar with respect to apocarotenoid content and gene expression profiles of 12 genes involved in apocarotenoid biosynthesis</p> <p>Carotenoid cleavage dioxygenase (CCD2) catalyzes the first step of crocin biosynthesis from carotenoid zeaxanthin and gets expressed at an extremely high level in the stigma as compared to corm, leaf, tepal, and stamen</p> <p>A C-class floral homeotic gene AGAMOUS (CsAG) gene is vital for stigma development of saffron. Its expression begins at yellow stage of stigma and increases sharply to orange stage, and continues to increase upto scarlet stage</p> <p>CsAP3 expression is maximum at late preanthesis of stigma development, while CsNAP expression increases abruptly at the scarlet stage of stigma</p> <p>CsNAP protein binds to the CArG1 region of CsAP3 promoter, and might be regulating CsAP3 expression indirectly by modulating CArG1 promoter</p>	[9–12, 14, 15, 18–27, 72, 92–95]

Table 1 (continued)

S. no.	Omics approach used	The gist of the main findings and outcomes	References
3.	Metabolomics	<p>Two novel saponins namely Azafrine 1 and Azafrine 2 have been isolated, purified, and structurally elucidated from the external part of saffron corm, suggesting that they may be acting as phytoprotectants</p> <p>¹H NMR-based metabolomics is useful to determine quality deterioration of saffron upon storage and for quality control</p> <p>Liquid chromatography coupled to electrospray ionisation time-of-flight mass spectrometry is an important tool for assessing saffron authenticity</p> <p>Tepals may have nutrition value owing to the presence of phytosterols and fatty acids, and can be processed as a source of flavonoids</p> <p>Metabolite profiling of stigma, tepal and stamen of <i>Crocus sativus</i> flower by ultra-performance liquid chromatography-quadrupole time-of-flight mass spectrometry (UPLC-QToF-MS/MS) has shown that coniferin and crocin-2 are special components in stigmas, while flavonoids are high in tepals</p> <p>High resolution mass spectrometry metabolomic studies in saffron from several countries has revealed that the phytochemical content varies among the samples of different countries</p> <p>At the yellow stage of stigma there are very low levels of crocetin, crocins, picrocrocin</p> <p>Picrocrocin and crocins are detected early in the orange stigma stage and increase rapidly in the red stigma stage</p> <p>The glycosylated products of crocetin reach maximum levels in the red stigma stage</p> <p>Saffron bioactive compounds are useful against coronary artery diseases, neurodegenerative disorders, bronchitis, asthma, diabetes, fever, colds, and metabolic syndrome</p> <p>Saffron can alleviate the symptoms of severe acute respiratory syndrome coronavirus 2 (COVID-19) patients and manage post-covid-19 syndrome</p> <p>The efficacy of saffron in managing depression is comparable to drugs like imipramine, fluoxetine, and citalopram</p> <p>Saffron can be used as an adjuvant in drug formulations as it acts as an immunity booster and anti-depressant</p>	[4, 33, 56, 63, 64, 96–101]
4.	Proteomics	<p>Thirty-six differentially accumulated proteins have been detected during somatic embryogenesis in <i>Crocus sativus</i> and involvement of ascorbate–glutathione cycle has been suspected in somatic embryo establishment</p> <p>Saffron protein database of stigma at different developmental stages is available through ProteomeXchange Consortium via the PRIDE partner repository with the data set identifier PXD009014 https://www.ebi.ac.uk/pride/archive/projects/PXD009014</p> <p>Two hundred and one differentially abundant protein species (DAPs) under cold stress affecting the floral initiation of saffron have been revealed using iTRAQ-based proteomics followed by real-time qPCR</p> <p>Saffron dormant corms exposed to low temperature stress do not bloom perhaps due to changes in the ‘reactive oxygen species–antioxidant system–starch/sugar interconversion homeostasis flowering pathway’</p>	[29, 30, 55]
5.	ABPP	<p>Drastic changes in the activity profile of cysteine proteases especially papain-like Cys proteases and vacuolar processing enzymes occur in the corms infected with <i>Fusarium oxysporum</i></p> <p>The activity of α-glycosidase AGLU1 gets suppressed upon <i>Fusarium oxysporum</i> infection in saffron corms irrespective of the F.o strain</p> <p>Activities of putative α-glycosidases (100-kD) and β-glucosidases (50–70 kD) increase upon <i>F. oxysporum</i> infection, while the activities of serine hydrolases (50, 60 kD) decrease</p> <p>Many β-glucosidases (45–60 kD) appear, while some (65–70 kD) disappear during <i>F. oxysporum</i> infection</p> <p>Glycosidase activity has a major role in maturation and development of stigma</p> <p>Sixty-seven active glycosidases that are differentially active during stigma development have been identified and quantified</p>	[29]

Table 1 (continued)

S. no.	Omics approach used	The gist of the main findings and outcomes	References
6.	miRNomics	Five miRNAs <i>csa-miR1</i> , <i>csa-miR2</i> , <i>csamiR3</i> , <i>miR414</i> and <i>miR837-5p</i> have been reported in <i>Crocus sativus</i> using in silico methods of EST analysis. These miRNAs may play roles in plant growth, disease resistance, senescence and stress responses	[57, 58]

severe acute respiratory syndrome coronavirus 2 (COVID-19) patients and managing the post-covid-19 syndrome [4]. Despite the importance of saffron in medicine and phytochemistry, modern approaches based on omics studies are relatively rare [3, 41, 42].

The metabolic and biochemical properties of saffron confirm its immense role in the pharmacognosy and pharma industry [4]. Studies on the binding potential of carotenoid pathway bioactive molecules for angiotensin-converting enzyme 2 (ACE2) receptor of SARS-CoV-2 show the possibility of using the saffron based remedy for novel coronavirus [66]. Flexible molecular docking followed by atomic level interaction study indicated that lutein and picrocrocin form various interactions with different amino acid residues of ACE2. In-depth analysis revealed that these interactions could be crucial for receptor-binding domain (RBD) binding and, therefore, potentially disrupting the interaction between RBD and ACE2. The study provides a clue for advanced studies involving in vitro, animal models and clinical studies. The efficacy of saffron in managing depression is comparable to drugs like imipramine, fluoxetine, and citalopram. The saffron metabolites can help manage stress and anxiety during the prolonged lockdown, isolation, and quarantine. Owing to all these beneficial properties and as an immunity booster, saffron extracts may be added in some drug formulations in future.

Bioinformatics for omics data analysis

Intricate regulatory networks of gene expression control the tissue and stage-specific accumulation of various metabolites. The systems biology approach integrates different omics technologies, including transcriptomics, proteomics, and metabolomics, so that biological systems are investigated in an integrated manner at different levels. The analysis of the complexity of the generated datasets needs to be integrated into the framework of known biological pathways. Bioinformatics plays a crucial role in data generation, analysis, and interpretation of the different omics technologies when mining meaningful information. It is crucial to interpret the massive amount of data generated through high throughput technologies and filter out useful information for interpretation by the researchers for comprehensive views on systems functionality [67]. Moreover, it provides resources

derived by exploiting -omics technologies or subsequent analyses, including sequence comparisons, gene family investigations, and molecular modelling, among others [68].

Omics-based technologies and other molecular research tools have led to the generation of a huge amount of information, necessitating bioinformatics advancement. It acts like a ‘feedback promotion’ and causes advancement in omics technologies due to its better handling of the ‘big data’. Bioinformatics creates and advances algorithms, computational techniques, and databases to better solve problems in the analysis of huge biological data. It has a key role in the textual mining of biological literature and query biological information. Bioinformatics tools can easily compare genetic and genomic data to better understand the evolutionary relationships between organisms. At a more integrative level, it analyzes the biological pathways and metabolic networks to give a better understanding into the systems biology. It helps in conducting simulation and modelling studies on DNA, RNA and proteins to understand their molecular interactions better, thus strengthening structural biology. It has assisted evolutionary biologists to (i) trace the evolution of organisms by calculating changes in their DNA; (ii) build complex models of populations for predicting the outcome; and (iii) share information about a large number of species [69]. The most relevant bioinformatics tools that have contributed to recent studies on saffron biology have been summarized in Table 2.

Large-scale expression profiling studies in saffron have generated huge amounts of data, and the discipline of bioinformatics has been indispensable for ‘deriving information’ from these data. As predicted [9], characterisation of the saffron stigmas through omics studies coupled with bioinformatics tools has generated vital novel information about the molecular basis of flavour, colour biogenesis, genomic organization, and the biology of *Crocus gynoecium* (Table 1).

GenoType and GenoDive are two important programs for analyzing genotypic diversity in clonal/asexual organisms [70]. The significance of genetic differentiation between accessions of saffron in Iran through the calculation of clonal diversity indices and Analysis of molecular variance (AMOVA) has been done using these tools [50]. Plant Intron and Exon Comparison (PIECE) is a comprehensive plant gene comparison and evolution database containing all the annotated genes described from 25 plant species with

Table 2 Bioinformatic tools and databases useful for omics data analysis

S. no.	Bioinformatic tools	Web address	Role	References
1.	SAM and BCF tools	https://www.htslib.org/ https://github.com/samtools/samtools https://www.htslib.org/ https://github.com/samtools/bcftools	Tools for processing and analysing sequencing data	[102]
2.	MEGA	http://www.megasoftware.net/	Comparative analysis and inferring evolutionary relationships of homologous sequences	[103–106]
3.	Trinity	https://github.com/trinityrnaseq/trinityrnaseq/releases/tag/v2.8.6	Tool for de novo transcriptome assembly of RNA-seq data	[87, 107]
4.	SMART 9	https://smart.embl.de/	Database for Identification and analysis of protein domains within protein sequences	[108]
5.	MPI bioinformatics toolkit	http://toolkit.tuebingen.mpg.de/	Web service for comprehensive and collaborative protein bioinformatic analysis	[109, 110]
6.	BiGGEsTS	http://kdbio.inesc-id.pt/software/biggests	Tool for revealing local coexpression of genes in specific intervals of time	[111]
7.	PlantGDB	http://www.plantgdb.org/	Database for comparative genomics/ genomic database encompassing sequence data for plants	[112]
8.	KEGG	http://www.kegg.jp/ http://www.genome.jp/kegg/	Database resource for biological interpretation of genome sequences and other high-throughput data	[80]
9.	TrichOME	http://www.plantrichome.org/	Comparative Omics database for plant trichomes	[113]
10.	PlantTFcat	https://www.zhaolab.org/PlantTFcat/	Tool for Identification and categorisation of plant transcription factors and transcriptional regulators	[114]
11.	Pln TFDB	http://plntfdb.bio.uni-potsdam.de/v3.0/	Database for functional and evolutionary study of plant transcription factors	[115, 116]
12.	Ensembl Plants	http://plants.ensembl.org	Database for visualising, mining and analysing plant genomic data	[117]
13.	Wego	http://wego.genomics.org.cn/	Web tool for plotting GO annotations	[118]
14.	edgeR	http://bioconductor.org/packages/edgeR/	Package for differential expression analysis of digital gene expression data	[119, 120]
15.	Bowtie	http://bowtie.cbcb.umd.edu/ https://sourceforge.net/projects/bowtie-bio/	Ultrafast, memory-efficient alignment program for aligning short DNA sequence reads to large genomes	[121]
16.	KaPPA-View	http://kpv.kazusa.or.jp/kpv4/	Web-based database for analysing omics data in plants	[122, 123]
17.	Transcriptogramer	http://bioconductor.org/packages/transcriptogramer	R package for transcriptional analysis based on protein–protein interaction	[124]
18.	Cufflinks	http://cole-trapnell-lab.github.io/cufflinks	Open-source software for RNA-Seq data analysis	[125, 126]
19.	Paintomics	http://www.paintomics.org/	Web based tool for joint visualization of transcriptomics and metabolomics data	[127]
20.	PIECE	https://probes.pw.usda.gov/piece/index.php	Database for plant gene structure comparison and evolution	[128]
21.	MISA-Web	http://misaweb.ipk-gatersleben.de/	Tool/web server for microsatellite prediction and counting	[129]
22.	Prodigal	https://github.com/hyattpd/Prodigal	Protein-coding gene prediction software tool	[83]
23.	GeneMarkS-T	http://topaz.gatech.edu/GeneMark/license_download.cgi	Tool for identification of protein-coding regions in RNA transcripts	[84]
24.	MaxQuant	https://maxquant.net/maxquant/	Quantitative proteomics software package for analysing large mass-spectrometric data sets	[86]
25.	Perseus	https://maxquant.net/perseus/	Software platform for interpreting protein quantification, interaction and post-translational modification data	[130]

Table 2 (continued)

S. no.	Bioinformatic tools	Web address	Role	References
26.	GenAlex	https://biology-assets.anu.edu.au/GenAlex/Welcome.html	Platform for population genetic analysis	[131]
27.	DnaSP	http://www.ub.edu/dnasp	Software package for DNA sequence polymorphism analysis of large data sets	[132]
28.	TransDecoder	https://github.com/TransDecoder/TransDecoder	Tool for Identification of potential coding regions within reconstructed transcripts	[82, 87]
29.	RepeatMasker package	https://www.repeatmasker.org/	Program to screen DNA sequences for interspersed repeats and low complexity DNA sequences	[133]
30.	GenoType and GenoDive	http://www.patrickmeirmans.com/software	Programs for the analysis of genetic diversity of asexual organisms	[70]
31.	psRNATarget	https://www.zhaolab.org/psRNATarget/	A small RNA target analysis server	[134]
32.	DESeq2 package	http://www.bioconductor.org/packages/release/bioc/html/DESeq2.html	Package for differential analysis of gene expression in plants	[135]
33.	Blast2Go	https://www.blast2go.com/	Platform for high-quality functional annotation and analysis of genomic datasets	[79]

available sequenced genomes, allowing a comparative analysis of saffron gene structures [71]. MicroSATellite (MISA) microsatellite finder is a tool for finding microsatellites in nucleotide sequences. Using this Perl script in saffron [72] has lead to the identification of microsatellites, also known as simple sequence repeats (SSRs), in *C. sativus*.

As *C. sativus* is a species without whole-genome sequencing, de novo transcriptome analysis provides an excellent and necessary platform to deepen the research on this plant at the molecular level [73]. Full-length reconstruction of transcriptomes from short-reads generated by Illumina sequencing technologies is the most challenging step in RNA-seq studies. In the absence of a reference genome, most common assembly strategies rely on Bruijn graph, including packages such as Trinity, SOAPdenovo-Trans, Velvet, Rnnotator and Oases [74]. Many studies have used Trinity for de novo assembly of saffron transcriptomics data [27, 73, 75], whereas others rely on strategies combining some of the aforementioned packages [15]. Alternative methods to Illumina sequencing, such as PacBio long-read sequencing, imply specialized software such SMRT Analysis software suite [26, 72, 76].

One of the main downstream applications after de novo assembly is transcript expression estimation which, in the absence of a reference genome, is performed by mapping reads against the assembled transcriptome. Algorithms that quantify expression from transcriptome mappings include RSEM, eXpress, Sailfish and kallisto, among others [74]. These algorithms typically depend on short read alignment programs such as Bowtie, which enables ultrafast and memory-efficient alignment of large sets of sequencing reads to a reference sequence [77]. In a recent study, the differentially expressed genes in saffron were identified via pair-wise comparisons of gene expression patterns between stigma and

the other four tissues (corm, leaf, tepal and stamen) by the ‘DESeq’ package [72]. The package is used for quantitative analysis of comparative RNA-seq data using shrinkage estimators for dispersion and fold change.

Functional annotation of the transcripts generated by the above-mentioned methods is typically achieved using similarity-detection tools such as BLAST [78]. Blast2GO has become a popular tool, allowing massive annotation of complete transcriptome datasets against a variety of databases, as well as GO functional classification and KEGG pathway enrichment [79]. Other software, WEGO (Web Gene Ontology Annotation Plot), allows visualizing, comparing and plotting GO annotation results. These tools have been widely used in the functional classification of uni-genes in RNA-seq studies on *Crocus sativus* [13, 20, 27, 32, 72]. Other annotation tools are based on the identification of specific domains in protein sequences. PlantTFcat is a high-performance web-based analysis tool that is designed to identify and categorise plant Transcription factor (TF)/Transcriptional regulator (TR)/Chromatin regulator (CR) genes from genome-scale protein and nucleic acid sequences by systematically analysing InterProScan domain patterns in protein sequences. Candidate transcription factors implicated in crocin biosynthesis in *Crocus sieberi* tepal and *C. sativus* stigma have been identified using PlantTFcat [13]. The Plant Transcription Factor Database (Pln TFDB), which is an integrative database that provides putatively complete sets of transcription factors and transcriptional regulators in plant species, has been used to identify genes encoding transcription factors in the network in saffron [57].

RNA-sequencing is a valuable tool to gain knowledge on high-level functions in biological systems. KEGG is an integrated database resource for biological interpretation of genome sequences and other high-throughput data [80]. It is

the reference knowledge base that integrates current knowledge on molecular interaction networks such as pathways and complexes (PATHWAY database), information about genes and proteins generated by genome projects (GENES/SSDB/KO databases) and information about biochemical compounds and reactions (COMPOUND/GLYCAN/REACTION databases) [81]. Hu [27] performed KEGG pathway analysis of differentially expressed genes (DEGs) and mapped 8251 unigenes into 130 standard pathways using KEGG database in saffron (*Crocus sativus* L.). Moreover, 14,671 genes were also annotated using KEGG database in Saffron [27].

TransDecoder identifies candidate coding regions within transcript sequences, such as those generated by de novo RNA-Seq transcript assembly using Trinity, or constructed based on RNA-Seq alignments to the genome using Tophat and Cufflinks [82]. It has been used in *Crocus sativus* L. protein domain annotation [29, 72]. Open reading frame detection and domain annotation from de novo assembled transcripts of *Crocus sativus* L. using TransDecoder along with two other algorithms (GeneMarkS-T, Prodigal) has led to the identification of 67 active glycosidases that are differentially active during stigma development, implying that glycosidase activity has a major role in the maturation of stigma [29]. Prodigal (PROkaryotic DYnamic programming Gene-finding Algorithm) is a fast, lightweight, open-source gene prediction program [83], while GeneMarkS-T is used for ab initio identification of protein-coding regions in RNA transcripts [84].

Different proteins that were either upregulated or downregulated in saffron under cadmium toxicity have been putatively identified using the MASCOT software search engine [85]. MaxQuant is a proteomics software for analysing large mass-spectrometric data sets [86]. It has been used for peptide and protein identification in different developmental stages of saffron stigma [29]. Peptide relative quantification between different MS runs was based solely on the LFQs, as calculated by MaxQuant (MaxLFQ algorithm). Another associated software platform (Perseus) supports researchers in the interpretation of protein quantification, interaction and post-translational modification, and is used for statistical analysis of MaxQuant outputs [86]. Saffron stigma spectra files submitted to an Andromeda search in MaxQuant were finally analysed and filtering of the results was done for post-translational modification, pattern recognition, time-series analysis in Perseus version 1.5.5.3. [29]. As discussed above, the identification and quantification of active glycosidases using ABPP could not have been possible without the support of bioinformatics tools [29]. Open Reading Frame Detection and Domain Annotation Softwares like Gene-MarkS-T [84], TransDecoder [87] and Prodigal [83] were used. Eight glycosidases (three GH3, three GH35, two GH116, and one GH1) were up-regulated more than

twofold in stage 4 stigmas. Moreover, the differential 110-kD β GH (Glucoside hydrolase) detected with labeling is most likely the GH116 enzyme CsTc017194, because this enzyme has a predicted molecular mass of 106 kD and is 4.5-fold up-regulated in stage 4 stigmas. The study illustrates the power of ABPP with bioinformatic predictive algorithms for quantitative glycosidase activity profiling on non-model plant species, like *Crocus sativus* L.

As shown by all these studies, there is immense scope for bioinformatics on elucidating biochemical functions of saffron proteins and its bioactive compounds [4, 29, 66].

Future prospects of saffron genetics

The studies cited above are a clear example of the tremendous advances in understanding the biology of saffron in recent years, yet there are many challenges and opportunities ahead. Deeper understanding of biological pathways has taken us closer to achieving the goal of developing a genome engineered *Crocus* sp. that can yield better quality saffron with high productivity. It will not be too far when these techniques enable editing genes encoding apocarotenoid biosynthesis through novel genome editing tools like CRISPR-Cas, making saffron breeding programs successful. In this context, whole-genome sequencing of *C. sativus* L. is becoming increasingly necessary to facilitate the exploitation of available gene-editing technologies.

Omics tools can be useful in locating sources of resistance and agronomically interesting traits for transfer to saffron by appropriate biotechnological tools [8] so that the resulting phenotype is able to better survive the challenges of erratic weather patterns and stresses of climate change. Such tools can also help appreciate the extent of the diversity of various geographic or genetic groups of cultivated saffron to infer relationships between groups and accessions. The information derived can be utilised for constructing biological pathways involved in the biosynthesis of principal components of saffron. Saffron metabolomics studies have revealed many peculiar properties of this interesting spice. However, the major challenge remains in identifying the incongruities in the biochemical pathways and the metabolic networks and correlating them with the phenotype.

Conclusion

Omics-based technologies have revolutionized biology, and saffron is no exception. These studies have helped to better understand the molecular mechanisms of flower development and growth, as well as the agents involved in apocarotenoid biosynthesis, its quality and diversity, pathology, and its potential as a therapeutic agent. Except for saffron

whole-genome sequencing, which is still awaited, a lot of useful information about saffron biology has been generated using omics-based techniques. These novel technologies helped discover new genes, study their expression, function, and evolutionary relationships, and made a plethora of information available to the scientific community, eventually taking us closer to developing a better quality high yielding saffron *Crocus* sp.

Acknowledgements AMH is grateful to National Mission on Himalayan Studies, Ministry of Environment, Forest and Climate Change, Government of India for funding saffron research in the form of a research grant to him. He is also thankful to the Scientists and staff of SKUAST-K for their help and support.

Funding Not applicable.

Data availability Not applicable.

Code availability Not applicable.

Declarations

Conflict of interest The authors have no conflict of interest and nothing to disclose.

Ethical approval This article does not contain any studies with human participants or animals performed by any of the authors.

Informed consent Not applicable.

References

- Kafi M et al (2018) An expensive spice saffron (*Crocus sativus* L.): a case study from Kashmir, Iran, and Turkey. In: Ozturk M et al (eds) Global perspectives on underutilized crops. Springer, Cham, pp 109–149
- Shahnoushi N et al (2020) Economic analysis of saffron production. In: Shahnoushi N et al (eds) Saffron. Elsevier, Amsterdam, pp 337–356
- Pandita D (2021) Saffron (*Crocus sativus* L.): phytochemistry, therapeutic significance and omics-based biology. In: Pandita D (ed) Medicinal and aromatic plants. Elsevier, Amsterdam, pp 325–396
- Husaini AM, Jan KN, Wani GA (2021) Saffron: a potential drug-supplement for severe acute respiratory syndrome coronavirus (COVID) management. *Heliyon* 7:e07068
- Husaini AM et al (2010) Saffron (*Crocus sativus Kashmirianus*) cultivation in Kashmir: practices and problems. *Funct Plant Sci Biotechnol* 4(2):108–115
- Ghaffari S (1986) Cytogenetic studies of cultivated *Crocus sativus* (Iridaceae). *Plant Syst Evol* 153(3):199–204
- Caiola MG (2005) Embryo origin and development in *Crocus sativus* L. (Iridaceae). *Plant Biosyst* 139(3):335–343
- Husaini AM (2014) Challenges of climate change: omics-based biology of saffron plants and organic agricultural biotechnology for sustainable saffron production. *GM Crops Food* 5(2):97–105
- Husaini AM et al (2009) Bioinformatics for saffron (*Crocus sativus* L.) improvement. *Commun Biometry Crop Sci* 4(1):3–8
- Husaini AM, Ashraf N (2010) Understanding Saffron biology using bioinformatics tools. *Saffron. Funct Plant Sci Biotechnol* 4:31–37
- D’Agostino N et al (2007) An EST database from saffron stigmas. *BMC Plant Biol* 7(1):1–8
- Frusciante S et al (2014) Novel carotenoid cleavage dioxygenase catalyzes the first dedicated step in saffron crocin biosynthesis. *Proc Natl Acad Sci* 111(33):12246–12251
- Ahrazem O et al (2018) Transcriptome analysis in tissue sectors with contrasting crocins accumulation provides novel insights into apocarotenoid biosynthesis and regulation during chromoplast biogenesis. *Sci Rep* 8(1):1–17
- Baba SA et al (2015) Comprehensive transcriptome analysis of *Crocus sativus* for discovery and expression of genes involved in apocarotenoid biosynthesis. *BMC Genomics* 16(1):1–14
- Jain M et al (2016) De novo transcriptome assembly and comprehensive expression profiling in *Crocus sativus* to gain insights into apocarotenoid biosynthesis. *Sci Rep* 6(1):1–13
- Ahrazem O et al (2015) Structural characterization of highly glucosylated crocins and regulation of their biosynthesis during flower development in *Crocus*. *Front Plant Sci* 6:971
- López AJ et al (2021) A new glycosyltransferase enzyme from Family 91, UGT91P3, is responsible for the final glucosylation step of crocins in saffron (*Crocus sativus* L.). *Int J Mol Sci* 22(16):8815
- Moraga ÁR et al (2009) Metabolite and target transcript analyses during *Crocus sativus* stigma development. *Phytochemistry* 70(8):1009–1016
- Castillo R, Fernández J-A, Gómez-Gómez L (2005) Implications of carotenoid biosynthetic genes in apocarotenoid formation during the stigma development of *Crocus sativus* and its closer relatives. *Plant Physiol* 139(2):674–689
- Ahrazem O et al (2019) Multi-species transcriptome analyses for the regulation of crocins biosynthesis in *Crocus*. *BMC Genomics* 20(1):1–15
- Nemati Z (2018) The origin of saffron
- Wafai AH et al (2017) Relative expression analysis of csag gene during different stages of stigma development in *Crocus sativus* L. (Saffron). *J Emerg Technol Innov Res* 4:272–274
- Gómez-Gómez L et al (2017) Unraveling massive crocins transport and accumulation through proteome and microscopy tools during the development of saffron stigma. *Int J Mol Sci* 18(1):76
- Wafai AH et al (2015) Comparative expression analysis of senescence gene CsNAP and B-class floral development gene CsAP3 during different stages of flower development in saffron (*Crocus sativus* L.). *Physiol Mol Biol Plants* 21(3):459–463
- Wafai AH, Husaini AM, Qadri RA (2019) Temporal expression of floral proteins interacting with CARG1 region of CsAP3 gene in *Crocus sativus* L. *Gene Rep* 16:100446
- Qian X et al (2019) Single-molecule real-time transcript sequencing identified flowering regulatory genes in *Crocus sativus*. *BMC Genomics* 20(1):1–18
- Hu J et al (2020) Transcriptome profiling of the flowering transition in saffron (*Crocus sativus* L.). *Sci Rep* 10(1):1–14
- Perez-Riverol Y et al (2019) The PRIDE database and related tools and resources in 2019: improving support for quantification data. *Nucleic Acids Res* 47(D1):D442–D450
- Husaini AM et al (2018) Multiplex fluorescent, activity-based protein profiling identifies active α -glycosidases and other hydrolases in plants. *Plant Physiol* 177(1):24–37
- Chen J et al (2021) Screening of key proteins affecting floral initiation of saffron under cold stress using iTRAQ-based proteomics. *Front Plant Sci* 12:708
- Morimoto K, van der Hoorn RA (2016) The increasing impact of activity-based protein profiling in plant science. *Plant Cell Physiol* 57(3):446–461

32. Ahrazem O et al (2010) The expression of a chromoplast-specific lycopene beta cyclase gene is involved in the high production of saffron's apocarotenoid precursors. *J Exp Bot* 61(1):105–119
33. Rubio A, Fernández JA, Gómez, L (2003) Biosynthesis of carotenoids in saffron. In: I International symposium on saffron biology and biotechnology, p 650
34. Carmona M et al (2007) A new approach to saffron aroma. *Crit Rev Food Sci Nutr* 47(2):145–159
35. Assimiadis MK, Tarantilis PA, Polissiou MG (1998) UV-Vis, FT-Raman, and 1H NMR spectroscopies of cis-trans carotenoids from saffron (*Crocus sativus* L.). *Appl Spectrosc* 52(4):519–522
36. Van Calsteren M-R et al (1997) Spectroscopic characterization of crocetin derivatives from *Crocus sativus* and *Gardenia jasminoides*. *J Agric Food Chem* 45(4):1055–1061
37. Knudsen JT, Tollsten L, Bergström LG (1993) Floral scents—a checklist of volatile compounds isolated by head-space techniques. *Phytochemistry* 33(2):253–280
38. Brighton CA (1977) Cytology of *Crocus sativus* and its allies (Iridaceae). *Plant Syst Evol* 128(3):137–157
39. Caiola MG, Caputo P, Zanier R (2004) RAPD analysis in *Crocus sativus* L. accessions and related *Crocus* species. *Biol Plant* 48(3):375–380
40. Alavi-Kia S et al (2008) Analysis of genetic diversity and phylogenetic relationships in *Crocus* genus of Iran using inter-retrotransposon amplified polymorphism. *Biotechnol Biotechnol Equip* 22(3):795–800
41. Alsayied NF et al (2015) Diversity and relationships of *Crocus sativus* and its relatives analysed by inter-retroelement amplified polymorphism (IRAP). *Ann Bot* 116(3):359–368
42. Larsen B et al (2015) Large intraspecific genetic variation within the Saffron-Crocus group (*Crocus* L., series *Crocus*; Iridaceae). *Plant Syst Evol* 301(1):425–437
43. Seberg O, Petersen G (2009) How many loci does it take to DNA barcode a crocus? *PLoS ONE* 4(2):e4598
44. Rubio-Moraga A et al (2009) Saffron is a monomorphic species as revealed by RAPD, ISSR and microsatellite analyses. *BMC Res Notes* 2(1):1–5
45. Moraga AR et al (2010) Intersimple sequence repeat markers for molecular characterization of *Crocus cartwrightianus* cv. *albus*. *Ind Crops Prod* 32(2):147–151
46. Imran S et al (2009) Studies in relation to molecular variability in saffron. In: III International symposium on saffron: forthcoming challenges in cultivation, research and economics, p 850
47. Nemati Z et al (2012) Isolation and characterization of a first set of polymorphic microsatellite markers in saffron, *Crocus sativus* (Iridaceae). *Am J Bot* 99(9):e340–e343
48. Keify F, Beiki AH (2012) Exploitation of random amplified polymorphic DNA (RAPD) and sequence-related amplified polymorphism (SRAP) markers for genetic diversity of saffron collection. *J Med Plants Res* 6(14):2761–2768
49. Seymour DK, Becker C (2017) The causes and consequences of DNA methylome variation in plants. *Curr Opin Plant Biol* 36:56–63
50. Busconi M et al (2015) AFLP and MS-AFLP analysis of the variation within saffron crocus (*Crocus sativus* L.) germplasm. *PLoS ONE* 10(4):e0123434
51. Busconi M et al (2018) Epigenetic stability in Saffron (*Crocus sativus* L.) accessions during four consecutive years of cultivation and vegetative propagation under open field conditions. *Plant Sci* 277:1–10
52. Busconi M et al (2021) Epigenetic variability among saffron crocus (*Crocus sativus* L.) accessions characterized by different phenotypes. *Front Plant Sci* 12:349
53. Schmidt T et al (2019) Adding color to a century-old enigma: multi-color chromosome identification unravels the autotriploid nature of saffron (*Crocus sativus*) as a hybrid of wild *Crocus cartwrightianus* cytotypes. *New Phytol* 222(4):1965–1980
54. Torricelli R et al (2019) Morphological and molecular characterization of Italian, Iranian and Spanish saffron (*Crocus sativus* L.) accessions. *Appl Ecol Environ Res* 17:1875–1887
55. Sharifi G et al (2012) Identification of differentially accumulated proteins associated with embryogenic and non-embryogenic calli in saffron (*Crocus sativus* L.). *Proteome Sci* 10(1):1–15
56. Bagri J et al (2017) Metabolic shift in sugars and amino acids regulates sprouting in Saffron corm. *Sci Rep* 7(1):1–10
57. Zinati Z, Shamloo-Dashtpajardi R, Behpouri A (2016) In silico identification of miRNAs and their target genes and analysis of gene co-expression network in saffron (*Crocus sativus* L.) stigma. *Mol Biol Res Commun* 5(4):233
58. Guleria P, Goswami D, Yadav KS (2012) Computational identification of miRNAs and their targets from *Crocus sativus* L. *Arch Biol Sci* 64(1):65–70
59. Chakraborty S (2016) Transcriptome from saffron (*Crocus sativus*) plants in Jammu and Kashmir reveals abundant soybean mosaic virus transcripts and several putative pathogen bacterial and fungal genera. *BioRxiv* 227:079186
60. Ambardar S et al (2016) Comparative metagenomics reveal phylum level temporal and spatial changes in mycobiome of belowground parts of *Crocus sativus*. *PLoS ONE* 11(9):e0163300
61. Cagliani LR et al (2015) NMR investigations for a quality assessment of Italian PDO saffron (*Crocus sativus* L.). *Food Control* 50:342–348
62. Petrakis EA et al (2015) Evaluation of saffron (*Crocus sativus* L.) adulteration with plant adulterants by 1H NMR metabolite fingerprinting. *Food Chem* 173:890–896
63. Ordoudi SA et al (2015) 1H NMR-based metabolomics of saffron reveals markers for its quality deterioration. *Food Res Int* 70:1–6
64. Vahedi M et al (2018) Quantitative HPLC-based metabolomics of some Iranian saffron (*Crocus sativus* L.) accessions. *Ind Crops Prod* 118:26–29
65. Tarantilis PA, Polissiou MG (1997) Isolation and identification of the aroma components from saffron (*Crocus sativus*). *J Agric Food Chem* 45(2):459–462
66. Ganai SA, Husaini AM (2021) Investigating binding potential of carotenoid pathway bioactive molecules for ACE2 receptor of SARS-CoV-2: possibility of a saffron based remedy for novel coronavirus! *J Horticult Postharvest Res* 4:69
67. Kumar A et al (2015) Systems biology for smart crops and agricultural innovation: filling the gaps between genotype and phenotype for complex traits linked with robust agricultural productivity and sustainability. *OMICS* 19(10):581–601
68. de Maturana EL et al (2019) Challenges in the integration of omics and non-omics data. *Genes* 10(3):238
69. Ogbé RJ, Ochalefu DO, Olaniru OB (2016) Bioinformatics advances in genomics—a review. *Int J Curr Res Rev* 8(10):05–11
70. Meirmans PG, Van Tienderen PH (2004) Genotype and genotype: two programs for the analysis of genetic diversity of asexual organisms. *Mol Ecol Notes* 4(4):792–794
71. Ahrazem O et al (2020) Differential interaction of Or proteins with the PSY enzymes in saffron. *Sci Rep* 10(1):1–11
72. Yue J et al (2020) Full-length transcriptome sequencing provides insights into the evolution of apocarotenoid biosynthesis in *Crocus sativus*. *Comput Struct Biotechnol J* 18:774–783
73. Tan H et al (2019) Transcriptome analysis reveals novel enzymes for apo-carotenoid biosynthesis in saffron and allows construction of a pathway for crocetin synthesis in yeast. *J Exp Bot* 70(18):4819–4834
74. Conesa A et al (2016) A survey of best practices for RNA-seq data analysis. *Genome Biol* 17:13

75. Mahmodi P et al (2014) Analysis of saffron stigma (*Crocus sativus* L.) transcriptome using SOAPdenovo and Trinity assembly software. *Crop Biotechnol* 4(6):35–46
76. Rhoads A, Au KF (2015) PacBio sequencing and its applications. *Genomics Proteomics Bioinform* 13(5):278–289
77. Langmead B (2010) Aligning short sequencing reads with Bowtie. *Curr Protoc Bioinform* 32(1):1–14
78. Altschul SF et al (1990) Basic local alignment search tool. *J Mol Biol* 215(3):403–410
79. Conesa A et al (2005) Blast2GO: a universal tool for annotation, visualization and analysis in functional genomics research. *Bioinformatics* 21(18):3674–3676
80. Kanehisa M et al (2016) KEGG as a reference resource for gene and protein annotation. *Nucleic Acids Res* 44(D1):D457–D462
81. Kanehisa M et al (2004) The KEGG resource for deciphering the genome. *Nucleic Acids Res* 32:D277–D280
82. Haas BJ, Papanicolaou A (2019) TransDecoder 5.5.0.
83. Hyatt D et al (2010) Prodigal: prokaryotic gene recognition and translation initiation site identification. *BMC Bioinform* 11(1):1–11
84. Tang S, Lomsadze A, Borodovsky M (2015) Identification of protein coding regions in RNA transcripts. *Nucleic Acids Res* 43(12):e78
85. Rao J, Lv W, Yang J (2017) Proteomic analysis of saffron (*Crocus sativus* L.) grown under conditions of cadmium toxicity. *Biosci J* 33(3):713
86. Tyanova S, Temu T, Cox J (2016) The MaxQuant computational platform for mass spectrometry-based shotgun proteomics. *Nat Protoc* 11(12):2301–2319
87. Haas BJ et al (2013) De novo transcript sequence reconstruction from RNA-seq using the Trinity platform for reference generation and analysis. *Nat Protoc* 8(8):1494–1512
88. Mir MA et al (2020) Deciphering genetic diversity analysis of saffron (*Crocus sativus* L.) using RAPD and ISSR markers. *Saudi J Biol Sci* 28:1308
89. Zheng H-J et al (2013) Construction of DNA finger printing for dry saffron. *Afr J Pharm Pharmacol* 7(43):2807–2812
90. Gedik A et al (2017) Genetic diversity of *Crocus sativus* and its close relative species analyzed by iPBS-retrotransposons. *Turk J Field Crops* 22(2):243–252
91. Zarini HN et al (2019) A comparative assessment of DNA fingerprinting assays of ISSR and RAPD markers for molecular diversity of Saffron and other *Crocus* spp. in Iran. *The Nucleus* 62(1):39–50
92. IqbalMzr J et al (2013) Relative expression of apocarotenoid biosynthetic genes in developing stigmas of *Crocus sativus* L. *J Crop Sci Biotechnol* 16(3):183–188
93. Diretto G et al (2019) UGT709G1: a novel uridine diphosphate glycosyltransferase involved in the biosynthesis of picrocrocin, the precursor of safranal in saffron (*Crocus sativus*). *New Phytol* 224(2):725–740
94. Zhou G et al (2020) Flower cultivation regimes affect apocarotenoid accumulation and gene expression during the development of saffron stigma. *Hortic Environ Biotechnol* 61(3):473–484
95. Gao G et al (2021) Transcriptomic analysis of saffron at different flowering stages using RNA sequencing uncovers cytochrome P450 genes involved in crocin biosynthesis. *Mol Biol Rep* 48(4):3451–3461
96. Rubio-Moraga Á et al (2011) Triterpenoid saponins from corms of *Crocus sativus*: localization, extraction and characterization. *Ind Crops Prod* 34(3):1401–1409
97. Guizarro-Díez M et al (2015) Metabolomic fingerprinting of saffron by LC/MS: novel authenticity markers. *Anal Bioanal Chem* 407(23):7197–7213
98. Consonni R et al (2016) On the traceability of commercial saffron samples using ¹H-NMR and FT-IR metabolomics. *Molecules* 21(3):286
99. Feizy J, Reyhani N (2016) Gas chromatographic determination of phytosterols and fatty acids profile in saffron petals. *Can Chem Trans* 4(3):389–397
100. Xu S et al (2019) Discrimination of different parts of saffron by metabolomic-based ultra-performance liquid chromatography coupled with high-definition mass spectrometry. *Chem Biodivers* 16(10):e1900363
101. Gikas E, Koulakiotis NS, Tsarbopoulos A (2021) Phytochemical differentiation of saffron (*Crocus sativus* L.) by high resolution mass spectrometry metabolomic studies. *Molecules* 26(8):2180
102. Danecek P et al (2021) Twelve years of SAMtools and BCFtools. *GigaScience* 10(2):008
103. Kumar S et al (2018) MEGA X: molecular evolutionary genetics analysis across computing platforms. *Mol Biol Evol* 35(6):1547–1549
104. Kumar S, Stecher G, Tamura K (2016) MEGA7: molecular evolutionary genetics analysis version 7.0 for bigger datasets. *Mol Biol Evol* 33(7):1870–1874
105. Tamura K, Stecher G, Kumar S (2021) MEGA11: molecular evolutionary genetics analysis version 11. *Mol Biol Evol* 38(7):3022–3027
106. Tamura K et al (2011) MEGA5: molecular evolutionary genetics analysis using maximum likelihood, evolutionary distance, and maximum parsimony methods. *Mol Biol Evol* 28(10):2731–2739
107. Grabherr MG et al (2011) Full-length transcriptome assembly from RNA-Seq data without a reference genome. *Nat Biotechnol* 29(7):644
108. Letunic I, Khedkar S, Bork P (2021) SMART: recent updates, new developments and status in 2020. *Nucleic Acids Res* 49(D1):D458–D460
109. Zimmermann L et al (2018) A completely reimplemented MPI bioinformatics toolkit with a new HHpred server at its core. *J Mol Biol* 430(15):2237–2243
110. Gabler F et al (2020) Protein sequence analysis using the MPI bioinformatics toolkit. *Curr Protoc Bioinform* 72(1):e108
111. Gonçalves JP, Madeira SC, Oliveira AL (2009) BiGGesTS: integrated environment for biclustering analysis of time series gene expression data. *BMC Res Notes* 2(1):1–11
112. Duvick J et al (2007) PlantGDB: a resource for comparative plant genomics. *Nucleic Acids Res* 36:D959–D965
113. Dai X et al (2010) TrichOME: a comparative omics database for plant trichomes. *Plant Physiol* 152(1):44–54
114. Dai X et al (2013) PlantTFcat: an online plant transcription factor and transcriptional regulator categorization and analysis tool. *BMC Bioinform* 14(1):1–6
115. Pérez-Rodríguez P et al (2010) PlnTFDB: updated content and new features of the plant transcription factor database. *Nucleic Acids Res* 38:D822–D827
116. Riaño-Pachón DM et al (2007) PlnTFDB: an integrative plant transcription factor database. *BMC Bioinform* 8(1):1–10
117. Bolser DM et al (2017) Ensembl plants: integrating tools for visualizing, mining, and analyzing plant genomic data. In: van Dijk ADJ (ed) *Plant genomics databases*. Springer, pp 1–31
118. Ye J et al (2006) WEGO: a web tool for plotting GO annotations. *Nucleic Acids Res* 34:W293–W297
119. Robinson MD, McCarthy DJ, Smyth GK (2010) edgeR: a Bioconductor package for differential expression analysis of digital gene expression data. *Bioinformatics* 26(1):139–140
120. McCarthy DJ, Chen Y, Smyth GK (2012) Differential expression analysis of multifactor RNA-Seq experiments with respect to biological variation. *Nucleic Acids Res* 40(10):4288–4297

121. Langmead B et al (2009) Ultrafast and memory-efficient alignment of short DNA sequences to the human genome. *Genome Biol* 10(3):1–10
122. Tokimatsu T et al (2006) KappA-View: a tool for integrating transcriptomic and metabolomic data on plant metabolic pathway maps. In: Saito K et al (eds) *Plant metabolomics*. Springer, pp 155–163
123. Tokimatsu T et al (2005) Kappa-View. A web-based analysis tool for integration of transcript and metabolite data on plant metabolic pathway maps. *Plant Physiol* 138(3):1289–1300
124. Morais DA, Almeida RM, Dalmolin RJ (2019) Transcrip-togramer: an R/Bioconductor package for transcriptional analysis based on protein–protein interaction. *Bioinformatics* 35(16):2875–2876
125. Trapnell C et al (2010) Transcript assembly and quantification by RNA-Seq reveals unannotated transcripts and isoform switching during cell differentiation. *Nat Biotechnol* 28(5):511–515
126. Trapnell C et al (2012) Differential gene and transcript expression analysis of RNA-seq experiments with TopHat and Cufflinks. *Nat Protoc* 7(3):562
127. García-Alcalde F et al (2011) Paintomics: a web based tool for the joint visualization of transcriptomics and metabolomics data. *Bioinformatics* 27(1):137–139
128. Wang Y et al (2013) PIECE: a database for plant gene structure comparison and evolution. *Nucleic Acids Res* 41(D1):D1159–D1166
129. Beier S et al (2017) MISA-web: a web server for microsatellite prediction. *Bioinformatics* 33(16):2583–2585
130. Tyanova S et al (2016) The Perseus computational platform for comprehensive analysis of (prote)omics data. *Nat Methods* 13(9):731–740
131. Peakall R, Smouse PE (2006) GENALEX 6: genetic analysis in Excel. Population genetic software for teaching and research. *Mol Ecol Notes* 6(1):288–295
132. Rozas J et al (2017) DnaSP 6: DNA sequence polymorphism analysis of large data sets. *Mol Biol Evol* 34(12):3299–3302
133. Smit A, Hubley R, Green P (2013) RepeatMasker. 2013. Institute for Systems Biology, Seattle
134. Dai X, Zhao PX (2011) psRNATarget: a plant small RNA target analysis server. *Nucleic Acids Res* 39s:W155–W159
135. Love MI, Huber W, Anders S (2014) Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. *Genome Biol* 15(12):1–21

Publisher's Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.