Understanding Saffron Biology using Bioinformatics Tools

Amjad M. Husaini 1 • Nasheeman Ashraf 2

1 Division of Plant Breeding and Genetics, Sher-e-Kashmir University of Agricultural Sciences and Technology of Kashmir, Jammu and Kashmir, 191121, India
2 Indian Institute of Integrative Medicine, Canal road, Jammu, Jammu and Kashmir, India

Corresponding author: * amjadhusaini@yahoo.com, dr.amjadhusaini@hotmail.com

ABSTRACT

Saffron (Crocus sativus L.) is a sterile triploid plant that belongs to the Iridaceae (Liliales, Monocots) whose genomes are relatively large and are poorly characterized. Among the 85 species belonging to the genus Crocus, saffron is the most fascinating and intriguing species. The word “saffron” is derived from the Arabic word “zafran”, which translates to “yellow”. Saffron was an important cultivated plant during the period of the Ottoman Empire, but its production has decreased with time (Arslan et al. 2007). Total world saffron production is estimated at 220,000 kg (220 metric tones), of which about 90% is produced in Khorasan Province, Iran (Jahan and Jahani 2007) and the remaining in Greece, Spain, Italy and India (Kashmir). Saffron introduction into new areas should be encouraged as it is a unique crop in terms of its potential and is recognized as red gold (Yadollahi et al. 2007). It is the highest priced spice in the world at around $500 kg−1 of saffron (Fernández 2007).

INTRODUCTION

Saffron (Crocus sativus L.) is a sterile triploid plant that is naturally propagated vegetatively by daughter corms developing on a mother corm. It is a member of the Iridaceae (Liliales, Monocots) whose genomes are relatively large and are poorly characterized (Fernández 2004). Among the 85 species belonging to the genus Crocus, saffron is the most fascinating and intriguing species. The word “saffron” is derived from the Arabic word “zafran”, which translates to “yellow”. Saffron was an important cultivated plant during the period of the Ottoman Empire, but its production has decreased with time (Arslan et al. 2007). Total world saffron production is estimated at 220,000 kg (220 metric tones), of which about 90% is produced in Khorasan Province, Iran (Jahan and Jahani 2007) and the remaining in Greece, Spain, Italy and India (Kashmir). Saffron introduction into new areas should be encouraged as it is a unique crop in terms of its potential and is recognized as red gold (Yadollahi et al. 2007). It is the highest priced spice in the world at around $500 kg−1 of saffron (Fernández 2007).

PRINCIPAL BIOACTIVE MOLECULES

Saffron, the dried red stigmas of C. sativus, is used as flavouring and colouring agent. Owing to extremely high demand from the dye, perfumery and flavouring industries, it is one of the most expensive spices on earth. A complex mixture of volatile and non-volatile compounds contributes to the overall aroma and flavour of saffron (Tarantilis and Polissiou 1997). For colour, the principal pigment is crocin, for smell the main component is safranal and for the special bitter flavour the main compound is the glycoside picrocrocin (Basker 1999). These compounds are derived from oxidative cleavage of the carotenoid zeaxanthin (Bouvier et al. 2003; Moraga et al. 2004). Recent studies have revealed a different volatile composition in unprocessed stigmas (Rubio et al. 2008, 2009). Recently bioinformatics tools for sequence homology allowed the identification and characterization of orthologs of carotenoid cleavage dioxygenase (CCD) involved in production of aroma from C. sativus (Bouvier et al. 2003; Rubio et al. 2008).

Saffron stigmas and corms are characterized by the presence of antifungal saponins (Hosseinzadeh and Younesi 2002) and a large number of defence proteins capable of binding to chitin and chitin oligosaccharides. In fact, a new class of defense chitinase namely Safchi A has recently been isolated from saffron (Castillo et al. 2007; Castillo and Gómez-Gómez 2009). Furthermore, different phenolic compounds (pyrogallic acid, kaempferol, p-coumaric acid and gallic acid) involved in stress responses have also been identified (Crungeo et al. 1986; Ebrahimzadeh et al. 1997). Peroxidase, catalase and superoxide dismutase activities have been detected in saffron corms in different develop-
mental stages (Keyhani and Keyhani 2004; Keyhani et al. 2006), and genomic approaches have enabled the identification of partial sequence homologues for these enzymes. Bioinformatics analysis of corm and stigma libraries from Crocus has identified several genes associated with defence responses (D’Agostino et al. 2007).

Apart from its use as a spice, saffron has been used for medicinal purposes as well (Schmidt et al. 2007). The evidence from the saffron pathway suggests that this is a potential therapy against a wide spectrum of tumors, such as leukemia, ovarian carcinoma, colon adenocarcinoma, rhabdomyosarcoma, papilloma, squamous cell carcinoma, and soft tissue sarcoma. In addition, saffron can be used to cure coronary heart disease and hepatitis, and to promote immunity. Most of these medicinal properties of saffron are due to crocin, safranal, picrocrocin and β-carotene (Abdullaev et al. 2003; Hosseinzadeh et al. 2004; Yeh et al. 2009; Dhar et al. 2009).

All these attributes make saffron an important crop for genome prospecting and demand approaches to build upon the knowledgebase of saffron in terms of genes, proteins and metabolites so as to provide researchers a platform for understanding the biogenesis of biologically important saffron metabolites. This would be a major step towards such biosynthetic pathways to enhance the accumulation of important metabolites and to provide clues to break the jinx of complex disease mechanisms by saffron metabolites.

**PRINCIPAL METABOLIC PATHWAYS AND GENES**

Carotenogenesis in ripening fruit and petals has been studied extensively (Hugueney et al. 1996; Hirschberg 2001; Moehs et al. 2001; Zhu et al. 2002, 2003; Kato et al. 2004). In flowers and fruits high concentration of carotenoids is correlated with upregulation of genes that enhance the flux of the biosynthetic pathway (Botella-Pavía and Rodríguez-Concepción 2006). In these tissues, development and carotenoid accumulation occurs alongside chloroplast to chromoplast transition while peculiarly in C. sativus, stigma development and carotenoid accumulation occurs concomitantly with the amyloplast to chromoplast transition and the stigma never turns green during this process (Grilli-Caiola and Canini 2004; Castillo et al. 2005). These carotenoids serve as precursors of physiologically important apocarotenoids of which crocetin represents an important metabolite for chemical and metabolic systems in saffron, identifying the functions of enzymatic genes is no longer satisfactory; instead we need to decipher the coordination and interaction among various metabolic pathways. In this context, other high throughput experiments, including genomics, transcriptomics, proteomics, and metabolomics provide us with the complementary information to elucidate such coordination. Here bioinformatics is needed for biological sequence analyses, transcriptome analyses, computational proteomics, computational metabolomics, bio-ontologies and biological databases. Thus bioinformatics can play a great role for data integration and analyses, and make sense out of different experimental data (Table 1).

**GENOMICS AND TRANSCRIPTOMICS**

Saffron is a triploid with karyotype 2n=3x=24 comprising of 8 triplids: triplets 1 and 2 include subacrocentric, triplets 3, 4 and 8 metaacentric, triplets 6 and 7 submetaacentric chromosomes and triplet 5 which shows an extreme difference in size of chromosomes. Genomes of several Crocus species have been investigated using a combination of genetic, molecular and cytological methods (Frello et al. 2000, 2004). More recently, many efforts have been made for a better understanding of the genomic organization of Crocus species. Seberg et al. (2009) presented the analysis of a proposed barcode sets (Chase et al. 2007) in the genus Crocus (Iridaceae) through the analysis of rpoC1, matK and tmH-psbA regions on 86 species of the genus Crocus and asserted the importance of barcoding as a promising technology.

Many marker-based studies have been performed for addressing the genetic diversity of Crocus. Moraga et al. (2009) analyzed the randomly amplified polymorphic DNA (RAPD) and inter simple sequence repeat (ISSR) marker profiles of 43 isolates of C. sativus to determine if this species is mono/polyploidy and assessed viability of saffron collected from different geographical areas. The results obtained from this study showed that all the clones collected from different geographical areas appeared as identical ones not only because of morphological characters but also at molecular level. In another study ISSR markers were used to characterize C. sativus and Crocus cartwrightianus, however, no differences were found thus confirming the earlier reports (Moraga et al. 2010). In contrast, molecular characterization based on RAPD markers revealed conside-
A reliable amount of genetic diversity among 10 elite saffron genotypes of Kashmir. The dendrogram based on molecular data divided the tested genotypes in two clusters at similarity coefficient of 44%, which showed a high level of genetic diversity between two clusters containing different genotypes (Imran et al. 2010). Alavi-Kia and colleagues (2008) used long terminal repeats (LTRs), a retrotransposon (RTN)-based marker study to analyze the genetic diversity and phylogenetic relationship in *Crocus* and to find the possible closest relatives of cultivated saffron from Iranian species of *Crocus*. Results of this study showed that except *C. sativus*, Iranian *Crocus* genus showed high diversity within and between species. In some cases, genetic variation was high among ecotypes of the same species from different geographical regions. These results also supported the possibility of Iranian *Crocus* species (*C. almehensis* and *C. michelso-nii*) as wild ancestors of saffron.

Apart from the whole genome sequencing projects, various efforts have been made for the generation of expressed sequence tags (ESTs) databases for different crops (Jantasuriyarat et al. 2005; Ramirez et al. 2005; Udall et al. 2006; Ashraf et al. 2009). ESTs provide an invaluable resource for analyzing gene expression associated with specific organs, growth conditions, developmental processes and responses to various environmental stresses and bridge the gap between genome sequence and gene function (Ashraf et al. 2009). In an effort to gain an understanding about the gene expression programs underlying accumulation of various apocarotenoids in *Crocus* stigmas, EST generation has been initiated recently (D’Agostino et al. 2007). The database is the first reference collection for the genomes of *Iridaceae*, for the molecular biology of stigma biogenesis and for the metabolic pathways underlying saffron secondary metabolism. D’Agostino et al. (2007) produced

Table 1: Plant bioinformatics databases useful for saffron analysis.

<table>
<thead>
<tr>
<th>Database</th>
<th>URL</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>Saffron genes</td>
<td><a href="http://www.saffrongenes.org">www.saffrongenes.org</a></td>
<td>EST collection from saffron stigmas</td>
</tr>
<tr>
<td>NCBI Plant</td>
<td><a href="http://www.ncbi.nlm.nih.gov/genomes/PLANTS/PlantList.html">www.ncbi.nlm.nih.gov/genomes/PLANTS/PlantList.html</a></td>
<td>Database on plant genomes</td>
</tr>
<tr>
<td>PlexDB</td>
<td><a href="http://www.plexdb.org">www.plexdb.org</a></td>
<td>Data on plant expression</td>
</tr>
<tr>
<td>GRIN</td>
<td><a href="http://www.ars-grin.gov/">www.ars-grin.gov/</a></td>
<td>Plant genetic resources</td>
</tr>
<tr>
<td>TAIR</td>
<td><a href="http://www.arabidopsis.org/">www.arabidopsis.org/</a></td>
<td>The Arabidopsis Information Resource</td>
</tr>
<tr>
<td>NASC</td>
<td><a href="http://arabidopsis.info/">http://arabidopsis.info/</a></td>
<td>Arabidopsis thaliana Information</td>
</tr>
<tr>
<td>MATDB</td>
<td><a href="http://maps.gsf.de/proj/thal/db/">http://maps.gsf.de/proj/thal/db/</a></td>
<td>Arabidopsis thaliana Information</td>
</tr>
<tr>
<td>NIAS db</td>
<td><a href="http://www.dna.afrc.go.jp/database/">www.dna.afrc.go.jp/database/</a></td>
<td>Database on rice</td>
</tr>
<tr>
<td>EXPASY</td>
<td><a href="http://www.expasy.org/links.html">www.expasy.org/links.html</a></td>
<td>Index to plant-specific databases</td>
</tr>
<tr>
<td>IRIS</td>
<td><a href="http://www.iris.irri.org">www.iris.irri.org</a></td>
<td>International Rice Information System</td>
</tr>
<tr>
<td>EMBL</td>
<td><a href="http://www.ebi.org">www.ebi.org</a></td>
<td>Sequence analysis package</td>
</tr>
<tr>
<td>OPEN-BIO</td>
<td><a href="http://www.open-bio.org">www.open-bio.org</a></td>
<td>Sequence Analysis package</td>
</tr>
<tr>
<td>GMOD</td>
<td><a href="http://www.gmod.org">www.gmod.org</a></td>
<td>Sequence Analysis package</td>
</tr>
<tr>
<td>TIGR</td>
<td><a href="http://www.tigr.org/software">www.tigr.org/software</a></td>
<td>Sequence Analysis package</td>
</tr>
<tr>
<td>SBML</td>
<td><a href="http://www.sbml.org">www.sbml.org</a></td>
<td>Metabolomics package</td>
</tr>
<tr>
<td>CROPFORGE</td>
<td><a href="http://www.cropforge.org">www.cropforge.org</a></td>
<td>Metabolomics package</td>
</tr>
<tr>
<td>ISCB</td>
<td><a href="http://www.iscb.org">www.iscb.org</a></td>
<td>International Society for Computational Biology</td>
</tr>
<tr>
<td>PRINTS</td>
<td><a href="http://www.bioinf.manchester.ac.uk/dbbrowser/PRINTS/index.php">www.bioinf.manchester.ac.uk/dbbrowser/PRINTS/index.php</a></td>
<td>Database of protein fingerprints</td>
</tr>
<tr>
<td>Phred/Phrap/Consed</td>
<td><a href="http://www.phrap.org">www.phrap.org</a></td>
<td>Programme for EST assembly</td>
</tr>
<tr>
<td>Arachne</td>
<td><a href="http://www.broad.mit.edu/wga/">www.broad.mit.edu/wga/</a></td>
<td>Tool for assembling genome sequence</td>
</tr>
<tr>
<td>GAP4</td>
<td><a href="http://staden.sourceforge.net/overview.html">http://staden.sourceforge.net/overview.html</a></td>
<td>Tool for sequence assembly</td>
</tr>
<tr>
<td>AMOS</td>
<td><a href="http://www.tigr.org/software/AMOS/">www.tigr.org/software/AMOS/</a></td>
<td>Whole-genome shotgun assembly</td>
</tr>
<tr>
<td>KEGG</td>
<td><a href="http://www.genome.jp/kegg/">www.genome.jp/kegg/</a></td>
<td>Pathway maps and pathway modules</td>
</tr>
<tr>
<td>Metacyc</td>
<td>metacyc.org/</td>
<td>Database of non-redundant, experimentally elucidated metabolic pathways</td>
</tr>
<tr>
<td>Geneontology</td>
<td><a href="http://www.geneontology.org">www.geneontology.org</a></td>
<td>Database for annotation of gene sequences</td>
</tr>
<tr>
<td>CAP3</td>
<td>pbl.univ-lyon1.fr/cap3.php</td>
<td>Sequence assembly programme</td>
</tr>
<tr>
<td>GeneSpring</td>
<td><a href="http://www.agilent.com/chem/geneSpring">www.agilent.com/chem/geneSpring</a></td>
<td>Analysis of microarray data</td>
</tr>
<tr>
<td>CArRAY</td>
<td><a href="http://caarray.nei.nih.gov">http://caarray.nei.nih.gov</a></td>
<td>Guides the annotation and exchange of array data</td>
</tr>
<tr>
<td>SWISS-2D PAGE</td>
<td><a href="http://au.expasy.org/ch2d">http://au.expasy.org/ch2d</a></td>
<td>Contains data on proteins identified on various 2-D PAGE and SDSPAGE reference maps</td>
</tr>
<tr>
<td>Melanie</td>
<td><a href="http://au.expasy.org/melanie">http://au.expasy.org/melanie</a></td>
<td>Two-dimensional electrophoresis (2-DE) gel analysis platform</td>
</tr>
<tr>
<td>Flicker</td>
<td><a href="http://open2dprot.sourceforge.net/Flicker">http://open2dprot.sourceforge.net/Flicker</a></td>
<td>An open-source stand-alone computer program for visually comparing 2D gel images</td>
</tr>
<tr>
<td>PEDRo</td>
<td><a href="http://pedro.man.ac.uk/">http://pedro.man.ac.uk/</a></td>
<td>Configurable data entry tool for XML</td>
</tr>
<tr>
<td>Emowse</td>
<td><a href="http://emboss.sourceforge.net/">http://emboss.sourceforge.net/</a></td>
<td>Open Source software analysis package specially developed for the needs of the molecular biology</td>
</tr>
<tr>
<td>Mascot</td>
<td><a href="http://www.matrixscience.com/">www.matrixscience.com/</a></td>
<td>A powerful search engine that uses mass spectrometry data to identify proteins from primary sequence databases</td>
</tr>
<tr>
<td>SEQUEST</td>
<td><a href="http://fields.scripps.edu/sequest">http://fields.scripps.edu/sequest</a></td>
<td>Correlates uninterpreted tandem mass spectra of peptides with amino acid sequences from protein and nucleotide databases</td>
</tr>
<tr>
<td>GOLM</td>
<td>csbdb.mpimp-golm.mpg.de/gmd.html</td>
<td>Database provides public access to custom mass spectra libraries, metabolite profiling experiments and other necessary information related to the field of metabolomics</td>
</tr>
<tr>
<td>MASSBANK</td>
<td><a href="http://www.massbank.jp/index.html?lang=en">www.massbank.jp/index.html?lang=en</a></td>
<td>Database of comprehensive, high-resolution mass spectra of metabolites</td>
</tr>
<tr>
<td>Reactome</td>
<td><a href="http://www.reactome.org/">www.reactome.org/</a></td>
<td>A curated knowledgebase of biological pathways</td>
</tr>
<tr>
<td>Cytoscape</td>
<td><a href="http://www.cytoscape.org/">www.cytoscape.org/</a></td>
<td>An open source bioinformatics software platform for visualizing molecular interaction networks and integrating these interactions with gene expression profiles and other state data</td>
</tr>
<tr>
<td>Visant</td>
<td>visant.bu.edu/</td>
<td>Tool for biological networks and pathways</td>
</tr>
</tbody>
</table>
6,603 high quality ESTs from a saffron stigma cDNA library and grouped these into 1,893 clusters, each corresponding to a different expressed gene. The complete set of raw EST sequences and their electropherograms are maintained in the Saffron Genes database [http://www.saffrongenes.org]. This allows users to investigate sequence qualities and EST structural features. The database structure consists of a main MySQL relational database and two satellite databases mysql and cDNA. Further, they assigned a preliminary function using BLASTX to the UniprotKB/Swiss-Prot database. Gene ontology terms were assigned to the transcripts and it was found that in the molecular function ontology class, transcripts with catalytic and hydrolase activity were most represented while in case of the biological function class, the vast majority of the GO assignments corresponded to transport category. The saffron genes with enzymatic function were also mapped to KEGG databases so as to know which metabolic pathway they are involved in.

The abundance of ESTs in a particular contig is an indicative of mRNA abundance of that particular gene in the stigma tissue. Considering this, D’Agostino et al. (2007) looked for the contigs that composed of more than 20 ESTs. The highly expressed contigs were short chain dehydrogenases, an abscisic acid accumulations. A few putative isoprenoid GTases, one of which could represent the still missing enzyme responsible for the glycosylation of picrocrocin. Several contigs also encoded for putative transcription factors and the most abundantly expressed were Myb-like protein with high similarity to LhMyb (from Lilium, GenBank accession BAB40790), Myb8 (from Gerbera) (Elmaa et al. 2003) and Myb305 (from Antirrhinum) (Jackson et al. 1991). All these transcription factors are highly expressed in flowers. This is a good beginning on a long way towards overall understanding of various metabolic pathways and their regulatory mechanisms that lead to the synthesis of biologically important metabolites in saffron. Thus bioinformatics helped in providing information regarding the probable function of the genes and the pathways they might have a role to play. Use of these genes for genetic engineering of such metabolic pathways would lead to enhanced levels of the important metabolites in genetically engineered saffron (Husaini et al. 2009).

In Crocus, the most valued metabolites are synthesised in stigma tissue and that too in developmental stage specific manner (Husaini et al., 2009). The transcriptome of saffron stigmas is vital for throwing light on the molecular basis of flavor, color biosynthesis, genomic organization and the biology of the gynoecium of spices in general and saffron in particular (Husaini et al. 2009). Some volatile and non volatile metabolites are also present in other Crocus tissues. This tissue and stage specific accumulation of various metabolites might be controlled by intrinsic regulatory networks of gene expression. Identifying these networks and the hierarchical relationship between them are vital to the understanding of biological systems. Moraga et al. (2009) used a combination of approaches to study the pattern of expression of several candidate genes and correlate that with volatile production and organoleptic characteristics of saffron during development. The pattern of accumulation of apocarotenoids in developing saffron stigmas was investigated by extracting stigmas corresponding to six different developing stages and analyzing the extracts by HPLC. The picrocrocin and crocin were detected in the early stages and increased rapidly during the following stages of development. In order to identify candidate genes encoding enzymes involved in volatile biosynthesis, in silico screening of the stigma cDNA database previously described (D’Agostino et al. 2007) was done and a comparison was drawn between the apocarotenoid content and the expression profiles. The results showed that during the development of C. sativus stigmas, the 1-deoxyxylulose 5-phosphate synthase (DXS), was expressed at all the stages while 3 hydroxy 3 methylglutaryl CoA reductase (HMGR) was expressed at low levels suggesting that DXS plays an important role in apocarotenoid accumulation. Also two putative terpene synthases were identified in the EST collection (TS1 and TS2) and each showed a different expression profile. The TS1 transcript was detected in all stages and the stigma gene expression levels of this transcript was undetectable during the early stages of stigma development and reaching a peak at preanthesis and anthesis thus suggesting the strong role of this enzyme in the biosynthesis of apocarotenoids. Similarly, transcript levels of two carotenoid biosynthesis genes CsPSY (phytoene synthase) and CsPDS (phytoene desaturase) increased in the red stage.

In another study (Castillo et al. 2005), accumulation of apocarotenoids was studied during stigma development followed by monitoring expression of some important genes of apocarotenoid biosynthetic pathway. It was observed that with the transition of yellow undeveloped to red developed stigmas, there was an accumulation of zeaxanthin accompanied by increased expression of phytoene synthase, phytoene desaturase and lycopene β cyclase. There was also a massive accumulation of carotene apocarotenoids was studied during stigma development and slightly increased as the stigma developed. The transcript level of TS2 was undetectable during the early stages of stigma development and reaching a peak at preanthesis and anthesis thus suggesting the strong role of this enzyme in the biosynthesis of apocarotenoids. Similarly, transcript levels of two carotenoid biosynthesis genes CsPSY (phytoene synthase) and CsPDS (phytoene desaturase) increased in the red stage.
on “Genetic Resources of Saffron and Allies (Crocus spp.): CROCUSBANK” to create, characterise and exploit a germplasm collection (bank) in Crocus species (Fernández 2007). This plant material can then be used in selection programmes and serve as sources of resistances to be transferred between saffron clones through appropriate breeding and biotechnological programmes. In realizing this objective, bioinformatics tools will be invaluable for locating the genes involved in flower colour and the orange stage. The apocarotenoid picrocrocin and crocetin are of special interest as the only compounds present at high levels at the end of the orange stage (Husaini et al. 2009). The removal of stamens and the hand separation of stigmas from saffron flowers are labour intensive and leads to the high cost of saffron stigmas (Tsafaris et al. 2004). It is desirable to have saffron flowers, which do not form stamens, or even have carpels in place of stamens, thus doubling saffron production in a single flower while lowering the production cost. As C-class MADS-box gene function is essential for both stamen and carpel formation, Tsafaris et al. (2005) characterized the expression of MADS-box genes in Crocus flowers using several molecular biology techniques, bioinformatics tools and database resources. Using a combination of conventional methods (5’- and 3’- RACE PCR) and new technologies (Rolling Circle Amplification RACE, and familyRCA-RACE) Tsafaris et al. (2009) conducted a detailed study on the regulatory mechanisms involved in flowering time and development of saffron. Such studies help in understanding and exploiting the molecular mechanisms that control flower development in crocus and in realization of the objective of producing flowers with carpels in place of stamens. Further, this knowledge can even be used in molecular medicine. Recently T and B-cell epitopes of Iranian C. sativus were mapped using bioinformatics tools and the predicted peptides were found useful for vaccine development (Hassan et al. 2008).

**METABOLOMICS**

Metabolomics is the analysis of the complete pool of small metabolites in a cell at any given time (Rhee et al. 2006). Metabolomics may prove to be particularly important in plants due to the proliferation of secondary metabolites. In a metabolite profiling experiment, metabolites are extracted from tissues, separated and analyzed in a high-throughput manner. Metabolic fingerprinting looks at a few metabolites to help differentiate samples according to their phenotype or biological relevance (Edwards and Batley 2004). Bioinformatics tools of metabolomics help in the identification and characterization of a broad range of metabolites through reference to quantitative biochemical analysis (Edwards and Batley 2004). The aspect GOLM metabolome database and the MASSBANK database have been extensively used. Various databases like KEGG, Reactome, MetaCyc and GO-ontology further help to know about the biochemical pathways in which the identified metabolites act and perform role. Biogenesis of the three major metabolites, crocetin glycosides, picrocrocin, and safranal, which are responsible for saffron colour, bitter taste and aroma, takes place from the oxidative cleavage of the carotenoid zeaxanthin (Bouvier et al. 2003). This step results in the formation of crocetin dialdehyde, and the final product crocetin glycoside (Moraga et al. 2004); and two identical β-ionone molecules, hydroxyl β-cyclocytral, which is thought to be converted by an UDP-glucosyltransferase to picrocrocin and safranal. Metabolomic studies of saffron enzymes involved in flavonoid glucosylation (flavonoid glucosyltransferases; Rubio et al. 2009) and carotenoid biosynthesis (lycopene β-cyclase, Abrazem et al. 2009) have the potential to reveal the dynamics of these pathways. Analysis of apocarotenoid content of Crocus at various stages of stigma development revealed that yellow young stigmas contained very low levels of crocetin, crocins, and picrocrocin and mainly contained unidentified compounds with maximum wavelengths around 250 nm that reached the highest levels in the orange stage. The apocarotenoid picrocrocin and crocins were detected early in the orange stage and increased rapidly during the following stages of stigma development and reaching at their maximum in the red stage. Also the glycosylated products of crocetin reached the highest levels in the red stage. The accumulation of crocins of higher glucose content agreed with the expression patterns observed for UGTc32, a glucosyltransferase enzyme involved in crocin and crocetin glucosylation in C. sativus stigmas (Rubio et al. 2004). The levels of picrocrocin began increasing at the end of the orange stage and reached the highest levels at anthesis, the stage of flower development when stigmas are collected for saffron preparation and which is characterized by high levels of the volatile safranal (Carmona et al. 2007).

In addition to the apocarotenoids, saffron also contains volatile compounds and it has been estimated that around 150 such compounds are present in saffron out of which only one third have been identified (Winterhalter and Straubinger 2000). Further, different and characteristic profiles of volatile compounds have been obtained for each developmental stage of saffron. For example, in the yellow stage, low levels of volatiles were produced and the fatty acid derivatives predominated while in the orange stage, carotenoid derivatives were detected in addition to the fatty acid derivatives. In the red stage, the volatiles derived from carotenoids accumulated to high levels, and β-cyclocitrinal, generated by the degradation of carotenoids reached its maximum levels suggesting that β-carotene contributes in this stage to the pool of crocin and crocetin. At the scarlet stage, right before anthesis, the volatile propanoic acid, 2-methyl-2,2-dimethyl-1-(2-hydroxy-1-methylethyl) propyl ester accumulated at high levels, but their levels decreased at anthesis, when monoterpenes and carotenoids reached their maximum levels. Interestingly, safranal was not among the main volatiles generated in the fresh tissue where high levels of picrocrocin were observed. In contrast, safranal has been considered as the major aromatic compound in saffron, comprising as much as 60-70% of the essential oil content (Alonso et al. 1996; Tarantilis and Poliissiou 1997), suggesting that this compound is most probably generated by picrocrocin degradation during the dehydration process of the stigma (Raina et al. 1996). Among the monoterpenes, linalool was emitted at high levels at anthesis, and is commonly described as responsible for fresh and floral odours (Dobson 1993; Knudsen et al. 1993). In the postanthesis stage, the fatty acid derived volatiles became the main volatile compounds together with the degradation product 4-oxoisophorone, suggesting that products of sesquenece may differ from those actively produced in mature stigmas, which would have to make the flowers less attractive for pollinators. This reduced attractiveness is hypothesized to encourage pollinators to visit the hand-separated stigmas and thus increase the reproductive success of the plant (Negre et al. 2003; Theis and Raguso 2005). Keeping this in view studies should be focussed towards the identification of various other saffron metabolites too which may also lead to the identification of new compounds having important medicinal properties and can be used as novel or improved phytotherapeutic agents.

**COMPARATIVE GENOMICS**

Comparative studies would help in understanding the complexity in various metabolic pathways including carotenoid biosynthesis in saffron. Carotenoid biosynthesis and its regulation have been studied in various plant species like Zea mays (Harjes et al. 2008), Daucus carota (Chlouault et al. 2008), tomato (Giuliano et al. 1993), etc. These plants are studied in much more detail and more genomic information is available for them as compared to Crocus. Therefore, we can compare the genomic information of Crocus with these crops and determine the differences in carotenoid synthesis. Many enzymes involved in carotenoid biosynthetic pathways belong to gene families (Gallagher 2004). How members of a gene family differ in their function would further help in understanding these complex pathways. This would throw light on how carotenoid biosynthesis is regulated in...
stage and tissue specific manner in Crocus, and give an overview of evolution of these genes and pathways.

The first International Symposium on Saffron Biology and Biotechnology, held in Albacete, Spain (2003) stressed the need for the creation of a bank of germplasm and gene banks in saffron. In 2005, a programme was launched by the European Commission for the conservation, characterisation, collection and utilisation of genetic resources in agrobiodiversity. AGRI GEN RES and BACC projects successfully concerted efforts and new genes and partners from 9 countries (CROCUSBANK) was constituted with two main goals (Fernández 2007; de los Mozos-Pascual et al. 2010a). The first being the collection and reproduction of saffron bulbs from all countries that cultivate saffron; and second, the collection of saffron allies for research into the taxonomy, evolution, genetics, physiology, ecology and agronomy of the genus. Currently the project is under progress and 384 accessions of saffron and wild crocuses are being preserved, multiplied and partially characterised in the Bank of Plant Germplasm of Cuenca (Spain) (de los Mozos-Pascual et al. 2010a). Preliminary characterisation of 50 saffron accessions from Azerbaijan, France, India, Iran, Italy, Morocco, New Zealand, Spain and Turkey has been done for characters related to phenology, floral morphology and saffron production. In addition preliminary DNA characterisation on 25 accessions has also been performed using PCR-based molecular markers (de los Mozos-Pascual et al. 2010b). Here bioinformatics tools can help in appreciation of the extent of the diversity of various geographic and/or genetic groups of cultivated saffron. Molecular-based trees can be constructed using software like CLUSTAL-W, MultAlign to infer relationships among groups and accessions (Husaini et al. 2009).

INTEGRATED APPROACH

The above mentioned approaches can help in exploring the biology of metabolite production in Crocus and their tissue specific accumulation. However, if used together, these approaches and technologies hold the potential of revolutionizing the biology of saffron. In saffron the overall metabolic process which leads to the synthesis of carotenoid compounds is a coordinated effort of many individual pathways. Further, the metabolite composition of saffron is different at different stages. Also other flower parts like sepals and petals contain a different set of metabolites. In this context, parallel analysis of transcript and metabolite profiling can reveal unexpected gene to metabolite networks by correlating the expression pattern of all genes with accumulating pattern of all metabolites (Urbanczyk-Wochniak et al. 2006). This would lead to the identification of new genes, unravelling of metabolic pathways and developing of correlation networks which will help to understand how different genes and their expression pattern affect the production and accumulation of various apocarotenoid compounds. This knowledge will be of immense use to remodel the carotenoid biosynthetic pathways so as to arrive at enhanced accumulation of various biologically active and important metabolites. Thus there are many bioinformatics software available for integration of such data and generation of correlation networks, for example, Cytoscape, Visant, etc.

Integration and structured interrogation of metabolome and transcriptome datasets would provide the basis for the integration of genome and phenome data. Linking gene expression, protein sequence and protein structure data will integrate genomics, transcriptomics and proteomics. Further integration of these metabolomic data would create the foundation for advanced knowledge-bases and help address a range of biological questions.

CONCLUSION

Saffron is genetically a rich crop as far as its use as spice and medicine is concerned. Not much effort has been made so far towards exploring its potential and its subsequent improvement. With growing need of this crop, the call for the day is integration of multiple biotechnological approaches and use of bioinformatics tools for analysing and integrating the data so as to acquire maximum output towards genome prospecting of this crop. While transcriptomics, proteomics and metabolomics would generate data, bioinformatics would provide glue for integration of these diverse areas so as to provide more comparative, connected and holistic view about the biology of saffron. This will further pave way to the strategy for improvement of this crop in various aspects.

REFERENCES


Integrating and structured interrogation of metabolome and transcriptome datasets would provide the basis for the integration of genome and phenome data. Linking gene expression, protein sequence and protein structure data will integrate genomics, transcriptomics and proteomics. Further integration of these metabolomic data would create the foundation for advanced knowledge-bases and help address a range of biological questions.

CONCLUSION

Saffron is genetically a rich crop as far as its use as spice and medicine is concerned. Not much effort has been made so far towards exploring its potential and its subsequent improvement. With growing need of this crop, the call for the day is integration of multiple biotechnological approaches and use of bioinformatics tools for analysing and integrating the data so as to acquire maximum output towards genome prospecting of this crop. While transcriptomics, proteomics and metabolomics would generate data, bioinformatics would provide glue for integration of these diverse areas so as to provide more comparative, connected and holistic view about the biology of saffron. This will further pave way to the strategy for improvement of this crop in various aspects.
the anciently diverged monocots and eudicots. Plant Physiology 133, 1831-1842
Fernández JA (2007) Genetic resources of saffron and allies (Crocus spp.). Acta Horticulturae 739, 167-185
Sey O, Lee KJ (2004) Post-translational modifications and their biological func-
tions: Proteomic analysis and systematic approaches. Journal of Biochemistry and Molecular Biology 37, 35-44
Tarantilis PA, Polissiou MG (1997) Isolation and identification of the aroma components from saffron (Crocus sativus). Journal of Agricultural and Food Chemistry 45, 459-462
