CBCS
http://agrobiol.sggw.waw.pl/cbcs

Communications in Biometry and Crop Science
Vol. 4, No. 1, 2009, pp. 3–8

International Journal of the Faculty of Agriculture and Biology,
Warsaw University of Life Sciences, Poland

REVIEW ARTICLE

Bioinformatics for saffron (*Crocus sativus* L.) improvement

Amjad M. Husaini[1*], Shafiq A. Wani[1], Parvez Sofi[1], Abdul G. Rather[1], Ghulam A. Parray[1], Asif B. Shikari[1], Javid I. Mir[2]

[1] Division of Plant Breeding and Genetics, Sher-e-Kashmir University of Agricultural Sciences and Technology of Kashmir, Jammu & Kashmir, India, 191121.
[2] Central Institute of Temperate Horticulture, Rangreth, Old Airport, Jammu & Kashmir, India, 190007.
* Corresponding author: Amjad M. Husaini, E-mail: amjadhusaini@yahoo.com

CITATION: Husaini, A.M., Wani, S.A., Sofi, P., Rather, A.G., Parray, G.A., Shikari, A.B., Mir, J.I. (2009). Bioinformatics for saffron (*Crocus sativus* L.) improvement. *Communications in Biometry and Crop Science* 4 (1), 3–8.

Received: 2 September 2008, Accepted: 8 January 2009, Published online: 19 February 2009

ABSTRACT

Saffron (*Crocus sativus* L.) is a sterile triploid plant and belongs to the Iridaceae (Liliales, Monocots). Its genome is of relatively large size and is poorly characterized. Bioinformatics can play an enormous technical role in the sequence-level structural characterization of saffron genomic DNA. Bioinformatics tools can also help in appreciating the extent of diversity of various geographic or genetic groups of cultivated saffron to infer relationships between groups and accessions. The characterization of the transcriptome of saffron stigmas is the most vital for throwing light on the molecular basis of flavor, color biogenesis, genomic organization and biology of gynoecium of saffron. The information derived can be utilized for constructing biological pathways involved in the biosynthesis of principal components of saffron i.e., crocin, crocetin, safranal, picrocrocin and safchiA.

Key Words: *bioinformatics; saffron; genomics; transcriptomics; proteomics; metabolomics; in silico.*

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 (Fernandez, 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 kilograms (220 metric tones), of

which about 90% is produced in Khorasan Province, Iran (Jahan and Jahani, 2007) and the remainder in Greece, Spain, Italy and India (Kashmir).

Owing to extremely high demand from the dye, perfumery and flavoring industries, it is one of the most expensive spices on earth. The components of the spice "saffron" are localized in the red stigmatic lobes of *C. sativus* flower and these are responsible for its distinct color, flavor and smell (Himeno and Sana, 1987). For color the principal pigment is crocin, for smell the main component is safranal and for the special bitter flavor the main compound is the glycoside picrocrocin (Baskar, 1999). These compounds are derived from oxidative cleavage of the carotenoid zeaxanthin (Bouvier et al. 2003; Moraga et al., 2004). Besides these, a new class of defense chitinase namely Safchi A has recently been isolated from saffron (Castillo et al., 2007). There has been little success however in enhancing the levels of these bioactive molecules of commercial importance.

A revolution in molecular biology, statistics and information technology has stimulated the merger of some advanced technologies for understanding the complex web of interactions linking individual components of a living cell to the integrated behavior of an entire organism (Bruskiewich et al., 2006). The marriage between these advanced technologies has given birth to the discipline of bioinformatics. As large-scale expression profiling experiments with saffron can generate huge amounts of data about the saffron transcriptome, the discipline of bioinformatics can be used to extract information from the data. Characterization of the transcriptome of saffron stigmas is vital for throwing light on the molecular basis of flavor, color biogenesis, genomic organization and the biology of the gynoecium of spices in general and saffron particularly. The information derived can then be utilized to construct biological pathways involved in the biosynthesis of the principal components of saffron i.e., crocin, crocetin, safranal, picrocrocin and safchiA.

FREELY AVAILABLE BIOINFORMATICS TOOLS

Plant genome analysis requires a broad range of protocols and algorithms. Many of these protocols and algorithms are freely available with bioinformatics tools. Some useful open-source sequence-analysis packages can be freely downloaded from web resources like EMBOSS (The European Molecular Biology Open Software Suite; www.emboss.org), The Open-Bio community (www.open-bio.org), GMOD (Generic Model Organism Database; www.gmod.org), TIGR software site (www. tigr.org/software), SBML (Systems Biology Markup Language; www.sbml.org), and Expasy Web site at the Swiss Institute of Bioinformatics (www.expasy.ch). Some important databases useful for saffron genome analysis are listed in Table 1.

PUBLIC SEQUENCE DATABASES

Modern crop research in saffron biotechnology is greatly facilitated by the presence of information on the internet in the form of international public sequence databases like Genbank at the National Center for Biotechnology Information (NCBI; www.ncbi.nlm.nih.gov), the European Molecular Biology Laboratory (EMBL) sequence database hosted by the European Bioinformatics Institute (EBI; www.ebi.ac.uk), and the DNA Data Bank of Japan (DDBJ; www.ddbj.nig.ac.jp). Although basic sequence data submitted to any of these three databases are automatically mirrored to the other two databases on a routine basis it is still worth the effort to explore independently each site for gathering specialized information.

PHYLOGENETIC ANALYSIS OF SAFFRON

The first International Symposium on Saffron Biology and Biotechnology, held in Albacete, Spain, stressed the need for the creation of a bank of germplasm and gene banks in saffron. A joint consortium with partners from 9 countries (CROCUSBANK) was constituted

with two main goals (Fernandez, 2007). First, 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. 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.

Table 1. Plant bioinformatics databases useful for saffron genome analysis (URLs last accessed on 5 Jan., 2009).

Database	URL	Description
Saffron genes	www.saffrongenes.org	EST collection from saffron stigmas
NCBI Plant	http://www.ncbi.nlm.nih.gov/genomes/PLANTS/PlantList.html	Database on plant genomes
PlexDB	www.plexdb.org/	Data on plant expression
GRIN	www.ars-grin.gov/	Plant genetic resources
TAIR	www.arabidopsis.org	The Arabidopsis Information Resource
NASC	http://arabidopsis.info/	Arabidopsis thaliana Information
MATDB	http://mips.gsf.de/proj/thal/db/	Database on rice
NIAS db	http://www.dna.affrc.go.jp/database/	
PlantCare	http://bioinformatics.psb.ugent.be/webtools/plantcare/html/	Database on cis-acting regulatory DNA elements in plants
EXPASY	www.expasy.org/links.html	Index to other plant-specific databases
IRIS	www.iris.irri.org	International Rice Information System
EMBOSS	www.emboss.org	Sequence Analysis package
OPEN-BIO	www.open-bio.org	Sequence Analysis package
GMOD	www.gmod.org	Sequence Analysis package
TIGR	www.tigr.org/software	Sequence Analysis package
SBML	www.sbml.org	Metabolomics package
CROPFORGE	www.cropforge.org	Metabolomics package
ISCB	www.iscb.org	International Society for Computational Biology
PRINTS	http://www.bioinf.manchester.ac.uk/dbbrowser/PRINTS/index.php	Database of protein fingerprints

TRANSCRIPTOMICS, TRANSGENOMICS AND GENE MINING

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 (Fernandez, 2007). Gene profiles from DNA microarray technology provide a snapshot of life that maps to a cross section of genetic activities controlled by thousands of genes simultaneously. Transcriptome analysis of saffron plants, subjected to different photoperiod and temperature regimes can throw light on to genes that get up or down-regulated. Variable temperature during corm dormancy and subsequent low temperatures appear to be effective factors in saffron flower initiation (Koocheki et al., 2007). Comparison of environmental and management practices for saffron in Iran (Khorasan) and India (Kashmir) throw light on some basic climatic and topographic differences between the two regions viz., humidity, altitude, rainfall, soil-type and irrigation. The main similarities being in time of planting, harvesting and low temperatures during the

growing season (Kafi and Showket, 2007). How these differences and similarities, translating into gene expression, can be known using DNA microarray technology and bioinformatics tools. This kind of huge data bases generated by physiological, agronomic and gene expression studies can be analyzed under *in silico* to find agronomically important candidate genes in saffron and to identify chemical agent(s) that simulate the effect of these variable factors, so saffron can be grown under controlled conditions. Furthermore, molecular biologists and biotechnologists can utilize the knowledge generated to specifically tailor saffron plants for new geographical areas such as the East Midlands of England (Yadollahi et al., 2007). This will involve development of novel traits and agriculturally relevant characteristics through changes in gene regulation. Software tools can be employed for *in silico* analysis of the impact of such molecular intervention i.e., introduction of a regulatory sequence or a transgene, for enhanced adaptation to new geographical areas.

Bioinformatics tools and DNA microarray technology can be useful in locating sources of resistance and agronomically interesting traits for transfer to saffron by appropriate biotechnological tools. The removal of stamens and the hand separation of stigmas from saffron flowers is labor intensive and leads to the high cost of saffron stigmas (Tsaftaris 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, Tsaftaris et al. (2005) recently characterized the expression of MADS-box genes in crocus flowers using several molecular biology techniques, bioinformatics tools and database resources. 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 *Crocus sativus* were mapped using bioinformatics tools and the predicted peptides were found useful for vaccine development (Hassan et al., 2008).

FUNCTIONAL GENOMICS OF SAFFRON

The primary motivation of functional genomics research in saffron is to narrow the list of candidate genes implicated in the biological processes involved in the production of flavoring compounds and stigma pigments, so that their expression can be enhanced using a transgenic approach and hence improve quality of the saffron stigma. Bioinformatics can play an enormous technical role in sequence-level structural characterization of saffron genomic DNA. Earlier, the only major information resource for modern genotyping and sequence characterization available to saffron biologists was the *Arabidopsis thaliana* (L.) Heynh. genome, published in 2000 (AGI, 2000). This information resource was further strengthened with the completion of the rice (*Oryza sativa* L.) genome project and the availability of the rice genome sequence (IRGSP, 2005). Even though several crop genome-sequencing projects are rapidly constructing a rich and diverse repository of information about plant DNA sequences, an important database for saffron has been designed recently to manage and explore the expressed sequence tags (ESTs) from saffron stigmas (Agostino et al., 2007). The database is the first reference collection for the genomics of Iridaceae, for the molecular biology of stigma biogenesis and for the metabolic pathways underlying saffron secondary metabolism (Agostino et al., 2007).

GENES EXPRESSED IN CROCUS STIGMAS

Agostino et al. (2007) produced 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 electopherograms are maintained in a database. This allows users to investigate sequence qualities and EST structural features.

Putative transcripts determined to be associated with enzymes are organized into classes and can be viewed in terms of enzyme assignments to metabolic pathways. This represents a straightforward way to investigate the properties of the stigma transcriptome, which contains a series of interesting sequences (putative sex determination genes, lipid and carotenoid metabolism enzymes, transcription factors), whose function can now be tested using *in vivo* or *in vitro* approaches.

A contig (from *contiguous*) is a set of overlapping DNA segments derived from a single genetic source and is used to deduce the original DNA sequence of that source. Several such contigs have been characterized in saffron genome, and based on the presence of tentative consensus sequences categorized into groups of putative function. The important ones include:

(1) Cl000944:1 encoding non-heme-β-carotene-hydroxylase, highly expressed in saffron stigmas (Castillo et al., 2005);

(2) Cl000627:1 encoding a putative glucosyltransferase, very similar to UGTCs2, which is able to glycosylate crocetin *in vitro* (Moraga et al., 2004);

(3,4) Cl001532:1 and Cl001032:1 encoding putative isoprenoid GTases, one of which could represent the still missing enzyme responsible for the glycosylation of picrocrocin;

(5) Cl000348:1 encoding a Myb-like protein with high similarity to LhMyb (from *Lilium*, GenBank accession BAB40790), Myb8 (from *Gerbera*) (Elomaa et al., 2003) and Myb305 (from *Antirrhinium*) (Jackson et al., 1991) and probably acting as a putative transcription factor.

Further, a large number of Cytochrome P450 sequences are expressed in saffron stigmas, some at very high levels (For details see Agostino et al., 2007).

CONCLUSIONS

Bioinformatics can assist in all stages of genotyping experiments in saffron, viz.: from raw data capture (e.g., gel image processing), documentation and storage to semi-automated analysis of raw data into inferences (i.e., germplasm fingerprinting and mapping, alignments of DNA variation to RNA and protein structures to elucidate functional variance, etc.). It can assist a saffron biologist in many ways, e.g.

Biological sequence alignment: To compare two DNA or protein sequences;

Proteomics: Which proteins are present in a cell and which proteins interact with each other;

Structural genomics: Determine the (3D) structure of proteins encoded by the saffron genome;

Comparative genomics: The function of target genes and DNA regions of saffron can be revealed by studying their parallels in other organisms;

Transcriptomics: Use of mRNA transcripts to determine which genes are turned on/off in a particular tissue type or organ (stigmatic lobes, corms etc.) and how external management practices and cultivation techniques such as irrigation, fertilization and hormone application can change this expression;

Phylogenetic analysis: Apply statistical algorithms to assess germplasm relatedness and biodiversity among different saffron clones;

Systems biology: For *in silico* analysis of the impact of molecular intervention on the biological system of genetically modified saffron, for better adaptation and enhanced production of commercially important compounds.

REFERENCES

AGI (The Arabidopsis Genome Initiative). (2000). Analysis of the genome sequence of the flowering plant *Arabidopsis thaliana*. *Nature* 408, 796–815.

Agostino, N.D., Pizzichini, D., Chiusano, M.L., Giuliano, G. (2007). An EST database from saffron stigmas. *BMC Plant Biology* 7, 53.

Arslan, N., Gurbuz, B., Ipek, A., Ozean, S., Sarihan, E. (2007). The effect of corm size and different harvesting times on saffron (*Crocus sativus* L.) regeneration. *Acta Horticulturae* 739, 113–117.

Baskar, D. (1999). Saffron chemistry. In: Negbi, M. (Ed.), *Saffron (Crocus sativus L.)*, Harward Academic Publishers, Amsterdam.

Bouvier, F., Suire, C., Mutterer, J., Camara, B. (2003). Oxidative remodeling of chromoplast carotenoids: Identification of the carotenoid dioxygenase CsCCD and CsZCD genes involved in *Crocus* secondary metabolite biogenesis. *Plant Cell* 15, 47–62.

Bruskiewich, R., Metz, T., McLaren, G. (2006). Bioinformatics and crop information systems in rice research. *International Rice Research Notes* 31, 5–12.

Castillo, R., Fernandez, J.A., Gomez-Gomez, L. (2005). Implications of carotenoid biosynthetic genes in apocarotenoid formation during the stigma development of *Crocus sativus* and its closer relatives. *Plant Physiology* 139, 674–689.

Castillo, R., Gomez, G., Fernandez, J.A. (2007). SafchiA is a new class of defence chitinase from saffron (*Crocus sativus* L.). *Acta Horticulturae* 739, 195–202.

Elomaa, P., Uimari, A., Mehto, M., Albert, V.A., Laitinen, R.A., Teeri, T.H. (2003). Activation of anthocyanin biosynthesis in *Gerbera hybrida* (Asteraceae) suggests conserved protein-protein and protein- promoter interactions between the anciently diverged monocots and eudicots. *Plant Physiology* 133, 1831–1842.

Fernandez, J.A. (2004). Biology, biotechnology and biomedicine of saffron. *Recent Research of Development in Plant Science* 2,127–159.

Fernandez, J.A. (2007). Genetic resources of saffron and allies (*Crocus* spp.). *Acta Horticulturae* 739, 167–185.

Hassan, M., Babak, S., Sasan, M. (2008). T and B-cell epitopes prediction of Iranian saffron (*Crocus sativus*) profilin by bioinformatics tools. *Protein & Peptide Letters* 15, 280–285.

Himeno, H., Sana, K. (1987). Synthesis of crocin, picrocrocin and safranal by saffron stigma-like structure proliferated in vitro. *Agricultural and Biological Chemistry* 51, 2395–2400.

IRGSP (International Rice Genome Sequencing Project). (2005). The map-based sequence of the rice genome. *Nature* 436, 793–800.

Jackson, D., Culianez-Macia, F., Prescott, A.G., Roberts, K., Martin, C. (1991). Expression patterns of myb genes from Antirrhinum flowers. Plant Cell 3, 115–125.

Jahan, M., Jahani, M. (2007). The effects of chemical and organic fertilizers on saffron flowering. *Acta Horticulturae* 739, 81–86.

Kafi, M., Showket, T. (2007). A comparative study of saffron agronomy and production systems of Khorasan (Iran) and Kashmir (India). *Acta Horticulturae* 739, 123–132.

Koocheki, A., Ganjeali, A., Abbassi, F. (2007). The effect of duration of incubation and photoperiod on corm and shoot characteristics of saffron plant (*Crocus sativus* L.). *Acta Horticulturae* 739, 61–70.

Moraga, A.R., Nohales, P.F., Perez, J.A., Gomez-Gomez, L. (2004). Glucosylation of the saffron apocarotenoid crocetin by a glucosyltransferase isolated from *Crocus sativus* stigmas. *Planta* 219, 955–966.

Tsaftaris, A.S., Pasentsis, K., Iliopoulos, I., Polidoros, A.N. (2004). Isolation of three homologous API-like MADS-box genes in *Crocus sativus* L. and characterization of their expression. *Plant Science* 166, 1235–1243.

Tsaftaris, A.S., Pasentsis, K., Polidoros, A.N. (2005). Isolation of a differentially spliced C-type flower specific AG-like MADS-box gene from *Crocus sativus* and characterization of its expression. *Biologia Plantarum* 49, 499–504.

Yadollahi, A., Azam-Ali, S., Cocking, E., Shojaei, Z.A. (2007). Possibility of growth and development of saffron in the UK. *Acta Horticulturae* 739, 139–149.