Annals of Botany **111**: 335–336, 2013
doi:10.1093/aob/mct020, available online at www.aob.oxfordjournals.org

ANNALS OF BOTANY
Founded 1887

VIEWPOINT

Use or abuse of bioinformatic tools: a response to Samach

Natalia Muñoz-Fambuena[1], Carlos Mesejo[1], María C. González-Mas[2], Eduardo Primo-Millo[2], Manuel Agustí[1] and Domingo J. Iglesias[2,*]

[1]*Instituto Agroforestal Mediterráneo, Universidad Politécnica de Valencia, E-46022 Valencia, Spain and* [2]*Departamento de Citricultura y Producción Vegetal, Instituto Valenciano de Investigaciones Agrarias, E-46113 Moncada, Valencia, Spain*
** For correspondence. E-mail iglesias_dom@gva.es*

Received: 4 June 2012 Returned for revision: 14 November 2012 Accepted: 14 January 2013 Published electronically: 6 February 2013

In a recent paper, we described for the first time the effects of fruit on the expression of putative homologues of genes involved in flowering pathways. It was our aim to provide insight into the molecular mechanisms underlying alternate bearing in citrus. However, a bioinformatics-based critique of our and other related papers has been given by Samach in the preceding Viewpoint article in this issue of *Annals of Botany*. The use of certain bioinformatic tools in a context of structural rather than functional genomics can cast doubts about the veracity of a large amount of data published in recent years. In this response, the contentions raised by Samach are analysed, and rebuttals of his criticisms are presented.

Key words: Candidate genes, citrus, EST, flowering, functional genomics, homology, orthology.

INTRODUCTION

Alternate bearing is a result of complex metabolic and molecular regulatory pathways in which flowering-related genes are strongly implicated. In a recent paper published in this journal we reported, for the first time, the effects of fruit on the expression of putative homologues of genes involved in flowering pathways. It was our aim to provide insight into the molecular mechanisms underlying alternate bearing in citrus (*Citrus clementina*) due to crop load. For detailed information, please refer to our original paper (Muñoz-Fambuena *et al.*, 2011).

The main conclusion derived from our study was that fruit modulates seasonal expression of flowering genes in alternate-bearing 'Moncada' mandarin: novel and original research in the context of fruit crop physiology (indeed, the publication was awarded a Graduate Prize by the journal). The study of the expression pattern of flowering-genes of *on* (fully loaded) and *off* (without fruit) trees revealed that mainly *CiFT* and also floral-identity genes (*CsAP1* and *CsLFY*) are highly involved in the process. No reference to choosing any new developmental regulator, as Samach (2013) generalizes, was included in our manuscript. Moreover, he claims that '*the vast amount of new sequence data provides us with great possibilities for giant leaps in our understanding*', and he globally criticizes some current, frequently used methodological approaches. It is certainly the case that in recent decades an astonishing advance in the development of new databases and bioinformatic tools and platforms has taken place (Rafalski, 2002). Nevertheless, current genome programs generate large amounts of data that require significant processing, storage and delivery to the research community (e.g. Vassilev *et al.*, 2005), not always resulting in a precise or complete annotation of genes and associated proteins – especially so if referred to EST collections. Previous publications have already warned about the current limitations of bioinformatics (see Rhee, 2005) and in this context it is well known that, as the volume of sequence data increases exponentially, the number of genes of unknown function – and, therefore, the large amount of associated uncertainties, gaps or mismatches – does as well (see Mwololo *et al.*, 2010). It is fairly easy to understand how this excess of information may be conveniently interpreted when introduced into functional research, and likewise conclusions may be exaggerated in scenarios of *in silico* gene prospecting. In our opinion, however, this potential lack of precision and the continuous improvement of bioinformatic procedures should not be exploited to spread doubts about the conclusions of functional studies that are just focused on the role of previously described genes.

In his Viewpoint article, Samach alerts the reader to the problems that arise when mixing new databases with old ones and their corresponding notations. In the following paragraphs we will give concise evidence as to why the conclusions we presented previously are as precise as possible in terms of functionality, the main goal of our research.

(1) To analyse the expression of those genes previously identified and sequenced in citrus based on an EST collection, the primers we used were selected for the specific amplification of the targets according to the current literature and the corresponding public annotation. Any new developmental regulator was studied. Just as an example to illustrate that Samach's assertions are not necessarily as precise as stated, we point out that when you BLAST the EST supposedly containing *CiFT* to public databases (i.e. NCBI) the unequivocal output is, precisely, *CiFT*. In addition, it is worth mentioning that in combining *in silico* the old with new and other particular non-published databases, Samach appears to ignore that the increase in the number

of genes of a given family is not a reason to invalidate the functionality of a target, previously-annotated gene.

(2) To analyse the expression of those genes not yet identified or annotated, a *consensus* sequence formed by all EST homologues (from the Citrus Functional Genomics Project; http://bioinfo.ibmcp.upv.es/genomics/cfgpDB/) to the target *Arabidopsis* gene was built (see Huang and Madan, 1999), and a specific primer was designed for amplification. This methodology (with its understandable limitations) is not new and in fact it is widely represented in the bibliography.

(3) In order to provide transparent information for the reader, we decided to include in our study, together with the sequence of the primer, a table with details about the notation of the EST with highest similitude to the target gene previously described (see point 1) or the *consensus* sequence designed (point 2). The fact that this sequence was or was not annotated in the referenced database is not so relevant from a functional perspective.

(4) Finally, the use of sequences from new, published databases (e.g. Aleza *et al.*, 2009) instead of being from previous ones would only affect the nomenclature and codes in our published table.

All things considered, we understand that the ultimate goal of functional genomics is to integrate large amounts of data so as to increase basic knowledge about key biological processes. Our study into the molecular mechanisms underlying alternate bearing in fruit trees is a good example – with the known limitations in the use of EST collections. And we hope that further studies supported with new data will confirm or improve our published results – which is, after all, the ultimate goal of scientific research.

LITERATURE CITED

Aleza P, Juarez J, Hernandez M, Pina J, Ollitrault P, Navarro L. 2009. Recovery and characterization of a *Citrus clementina* Hort. ex Tan. 'Clemenules' haploid plant selected to establish the reference whole Citrus genome sequence. *BMC Plant Biology* **9**: 110. http://dx.doi.org/10.1186/1471-2229-9-110.

Huang X, Madan A. 1999. CAP3: a DNA sequence assembly program. *Genome Research* **9**: 868–877.

Muñoz-Fambuena N, Mesejo C, Gonzalez-Mas MC, Primo-Millo E, Agusti M, Iglesias DJ. 2011. Fruit regulates seasonal expression of flowering genes in alternate-bearing 'Moncada' mandarin. *Annals of Botany* **108**: 511–519.

Mwololo JK, Munyua JK, Muturi PW, Munyiri SW. 2010. An overview of advances in bioinformatics and its application in functional genomics. *Journal of Animal and Plant Sciences* **6**: 645–652.

Rafalski JA. 2002. Plant genomics: present state and a perspective on future developments. *Briefings in Functional Genomics and Proteomics* **1**: 80–94.

Rhee SY. 2005. Bioinformatics. Current limitations and insights for the future. *Plant Physiology* **138**: 569–570.

Samach A. 2013. Congratulations, you have been carefully chosen to represent an important developmental regulator! *Annals of Botany* **111**: 329–333.

Vassilev D, Leunissen J, Atanassov A, Nenov A, Dimov G. 2005. Application of bioinformatics in plant breeding. *Biotechnology and Biotechnological Equipment* **19**: 139–152.