

Identification of Distinctive Variants of the Olive Pollen Allergen Ole e 5 (Cu,Zn Superoxide Dismutase) throughout the Analysis of the Olive Pollen Transcriptome

Adoración Zafra¹, Rosario Carmona^{1,2}, José C. Jimenez-Lopez¹, Amada Pulido³,
M. Gonzalo Claros², and Juan de Dios Alché¹

¹ Estación Experimental del Zaidín (CSIC), Granada, Spain

{dori.zafra, juandedios.alche, josecarlos.jimenez}@eez.csic.es

² Departamento de Biología Molecular y Bioquímica, Universidad de Málaga, Málaga, Spain
{rosariocarmona, claros}@uma.es

³ Departamento de Fisiología Vegetal, Universidad de Granada, Granada, Spain
amadapulido@ugr.es

Abstract. Ole e 5 is an olive pollen allergen displaying high identity with Cu,Zn superoxide dismutases. Previous studies characterized biochemical variability in this allergen, which may be of relevance for allergy diagnosis and therapy. The generation of an olive pollen transcriptome allowed us to identify eight Ole e 5 sequences, one of them including a 24 nt deletion. Further *in silico* analysis permitted designing primers for PCR amplification and cloning from both cDNA and gDNA. A large number of sequences were retrieved, which experimentally validated the predictive NGS sequences, including the deleted enzyme. The PCR-obtained sequences were used for further scrutiny, including sequence alignment and phylogenetic analysis. Two model sequences (a complete sequence and a deleted one) were used to perform 3-D modeling and a prediction of the T- and B-cell epitopes. These predictions interestingly foresee relevant differences in the antigenicity/allergenicity of both molecules. Clinical relevance of differences is discussed.

Keywords: allergen, deletion, Ole e 5, pollen, Superoxide Dismutase.

1 Introduction

Olive pollen is one of the most important causes of respiratory allergy in the Mediterranean area [1]. To date, twelve allergens have been identified in olive pollen, and one in olive mesocarp, named as Ole e 1 to Ole e 13 [2,3]. Several studies have shown that olive pollen allergens possess a relatively high level of microheterogeneity in their sequence, which has been extensively characterized for allergens Ole e 1, Ole e 2 and

Ole e 7 [3,4,5]. Moreover, olive pollen contains a complex mix of allergen forms, which is highly dependent of the genetic background (olive cultivar) used for the analysis [4].

Allergen polymorphism is a relevant issue concerning allergy diagnosis and treatment. Heterogeneity of vaccines used for specific immunotherapy (SIT) is considered a challenging factor their efficacy or even worse, it may represent a cause for secondary sensitization of patients or the development of undesirable reactions. Therefore, accurate identification and quantification of the different allergenic forms present in the extracts and its reactivity is a major concern nowadays.

Ole e 5 is considered a minor allergen in olive pollen, with prevalence around 35%, which has been identified as a Cu,Zn superoxide dismutase (SOD). Due to its ubiquity, it is considered as a cross-reactive allergen in the pollen-latex-fruit syndrome [6]. Studies carried out over Ole e 5 show that it is a 16 kDa protein with a high identity to Cu,Zn SODs from other species. In olive pollen, the presence of up to five isoforms has been detected [4, 7-9]. Concerning the biological function of this protein in pollen, SODs catalyse the dismutation of the superoxide anions into molecular oxygen and hydrogen peroxide, therefore acting as an antioxidant able to remove reactive oxygen species (ROS) [10]. ROS are produced in both unstressed and stressed cells. Under unstressed conditions, the formation and removal of O₂ are in or close to balance. Within a cell, the SODs constitute the first line of defence against ROS [11] and they have been localized in cytosol, chloroplasts, peroxisomes and the apoplast [12,13]. ROS may also act as signalling molecules. Thus, the interaction pollen-pistil is mediated by the accumulation of ROS in the stigma, where Ole e 5, is considered one of the main actors [14,15]. Under stress conditions, a clear rise of Cu,Zn-SOD has also been detected [16].

Actual olive tree transcriptomes mainly rely on libraries from vegetative tissues using different NGS strategies. The peculiarity of the reproductive tissues, and the widely reported presence of numerous tissue-specific transcripts in pollen grains, made us to attempt a similar approach in the later. Hence, reproductive cDNA libraries were prepared to be sequenced using 454/Roche Titanium+ platform. This information is highly valuable to elucidate the mechanisms governing the allergy process, as well as its implication in the olive pollen metabolism through germination, stigma receptivity, and the interaction pollen-stigma.

2 Materials and Methods

2.1 Construction of the Olive Pollen Transcriptome

For the construction of the pollen transcriptome, mature pollen obtained from dehiscent anthers of the olive cultivar 'Picual' was used. The samples were thoroughly grinded with a pistil and liquid N₂ followed by the extraction of the total RNA as RNeasy Plant Mini Kit (Qiagen) manual instructions recommends. RNA integrity was checked by formaldehyde gel analysis [17]. The mRNAs were purified using the Oligotex mRNA mini kit (Qiagen). The concentration and quality of the mRNAs were determined by the Ribogreen method (Quant-it RiboGreen RNA Reagent and kit) and

the Agilent RNA 6000 Pico assay chip (Bioanalyzer 2100). The isolated mRNAs were subjected to 454/Roche Titanium+ sequencing.

2.2 Pre-processing

All reads obtained were pre-processed using the SeqTrimNext pipeline, in order to remove unsuitable reads, such as low quality, low complexity, linkers, adapters, vector fragments, polyA/polyT tails and contaminated sequences [18].

2.3 Assembling

The assembling strategy relies on the combination of two completely different algorithms to compensate assembling biases. After several assembling attempts, the best assembler combination was MIRA3 (based on overlap-layout-consensus algorithm) with Euler-SR (based on de Bruijn graphs) and a final reconciliation with CAP3.

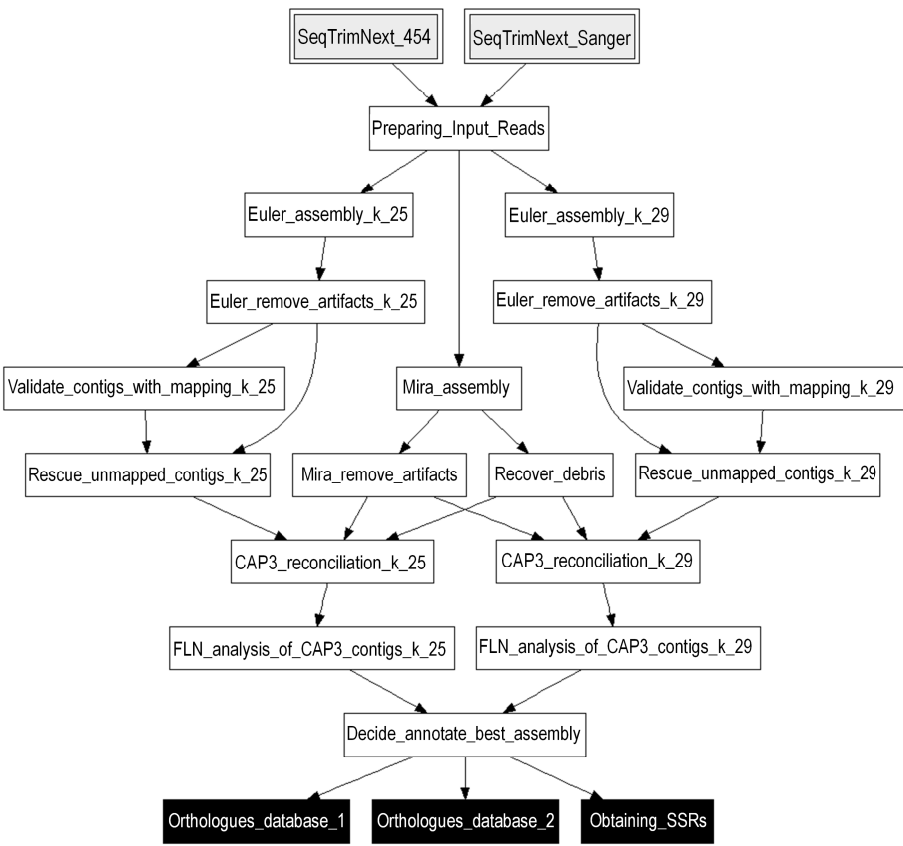


Fig. 1. Pipeline for the assembling of transcripts

2.4 Functional Annotation

A preliminary unigene analysis was performed using Full-LengtherNext. It provided a gene description, full-length unigenes, putative start and stop codons, the putative protein sequence, putative ncRNAs, an unknown unigenes. This gives a quick preview of the olive pollen transcriptome content. Full annotation was carried out using Sma3 to provide another gene description.

2.5 Selection of Cu,Zn-SOD Transcripts

Different strategies were defined to select such transcripts. First, EC codes of *A. thaliana* orthologous to olive sequences were extracted from annotations and subjected to analysis into the *Arabidopsis* reactome database of plant biological pathways. Finally, annotations were manually screened for specific enzymes selected from well-established bibliography resources.

2.6 Cloning of full Cu,Zn-SODs Sequences on the Basis of the Retrieved Sequences

Genomic DNA was obtained from the olive leaves of 7 cultivars and the extraction was carried out with the REDExtract-N-Amp Plant PCR (Sigma). Total RNAs corresponding to olive pollen from 17 olive cultivars were obtained using the RNeasy Plant Mini kit (Qiagen) and subjected to reverse transcription using the SuperScript II reverse transcriptase kit (Invitrogen). The gDNAs and cDNAs obtained were amplified with the following primers:

5' ATG GTG AAG GCC GTA ACA GTC 3' (forward) and
5' TCA ACC CTG AAG GCC AAT G 3' (reverse).

The PCR products were analyzed electrophoretically, and the amplified bands were excised, and purified with the MBL-Agarose Quikclean kit (Dominion). The purified PCR products were cloned in pGEM-Teasy vector system (Promega). *E. coli* DH5 α were transformed with the vector followed by a selection of the colonies: the plasmidic DNAs were extracted as indicated in the Real mini-prep turbo kit (Real) manual instruction. After double-checking the correct insertion of the fragment, it was sequenced in the Sequencing Service of the Institute of Parasitology and Biomedicine "López-Neyra" (IPBLN-CSIC, Granada, Spain). At least 3 clones of each cultivar were sequenced.

2.7 Alignment and *in silico* Analysis of the Sequences

The alignment of all the nucleotidic sequences obtained was performed by using the Clustal W software (<http://www.ebi.ac.uk/Tools/clustalw/>). The WinGene 1.0 software (Henning 1999) was used to generate the translation to aminoacidic sequences, which were aligned using the Clustal W software as well. The ScanProsite software

(<http://www.expasy.org/tools/scanprosite>) was used to identify potential post-translational modifications and SOD consensus sequences.

2.8 Phylogenetic Tree

Phylogenetic trees were constructed with the aid of the software Seaview [19] using the maximum likelihood (PhyML) method implemented with the JTT model of the most probable amino acid substitution calculated by the ProtTest 2.4 server [20]. The branch support was estimated by bootstrap resampling with 100 replications.

2.9 3D-Modeling

The two most distinctive forms of Cu,Zn SODs were subjected to 3D reconstruction (<http://swissmodel.expasy.org/workspace/>)[21,22,23,24] by using the 2q21 template (annotated as a dimer) available as PDB by means of the DeepView v3.7 software.

2.10 Identification of B- and T-cell Epitopes

To determine linear B- and T-cell, as well as conformational B-cell epitopes we made use of a set of tools as described for other olive pollen allergens like Ole e 12 [25].

3 Results

3.1 Screening of the Olive Pollen Transcriptome for Cu,Zn SOD Sequences

A total of eight inputs corresponding to Cu,Zn SOD were obtained from the database generated (Table1). The sequences were named OePOlee5-1 to OePOlee5-8 (*Olea europaea* Pollen allergen Ole e 5-1 to 8).

Table 1. Selected sequences from the olive pollen transcriptome identified as Cu,Zn SOD. Lengths from the sequences obtained after the assembly and from the codificant protein are indicated, as well as the missing aminoacids. Putative cellular localization is also shown.

Identifier	Nucleotides (NGS)	Aa (no UTRs)	Seq. description	Subcellular location
OePOlee5-1	677	128	Lacking N-terminal(24A)	Cytosol
OePOlee5-2	532	150	Lacking N-terminal (2aa)	Cytosol
OePOlee5-3	450	144	Deleted sequence (8 aa)	Cytosol
OePOlee5-4	621	152	Complete sequence	Cytosol
OePOlee5-5	587	192	Lacking C-terminal (36aa)	Chloroplast
OePOlee5-6	615	155	Complete sequence	Cytosol
OePOlee5-7	407	39	Lacking N-terminal (189aa)	Chloroplast
OePOlee5-8	758	152	Complete sequence	Cytosol

3.2 Alignment of the NGS-Retrieved Sequences of Cu,Zn SODs to the GenBank Database and Phylogenetic Analysis

BLAST query of the individual sequences against the GenBank database confirmed all of them as Cu,Zn SODs, either highly homologous to Cu,Zn SODs previously described in olive (Sequences OePOlee5-1,3,4,6, and 8), or in other tree species (sequences OePOlee5-2,5, and 7). Sequences alignment allowed the detection of microheterogeneities among the sequences, affecting several positions, however a high level of identity was the main characteristic observed (data not shown). The most distinctive feature observed was the presence of one relevant deletion (24 nucleotides) in one of the sequences (OePOlee5-3). The phylogenetic tree generated by including most relevant BLAST scores showed the presence of two differentiated clusters, putatively corresponding to those previously described as chloroplastidic and cytosolic proteins, respectively (Figure 2). However, not all the sequences used for the phylogeny were annotated as chloroplastidic or cytosolic by the describing authors.

3.3 Alignment of the Cloned Sequences from Different Olive Cultivars

The nucleotide sequences experimentally obtained from cDNA (51 sequences) and gDNA (21 sequences) were aligned separately (not shown). In both cases, most of the sequences were 456 nt long. Three exceptions were found where the length of the sequences were 432 nt in all cases, with a deletion of 24 nt positioned between the nt 252-276. These three exceptions corresponded to clones obtained from different olive cultivars and origins ('Arbequina' and 'Empeltre' from cDNA; 'Loaime' from gDNA).

All 72 aminoacidic sequences were aligned in order to identify postraslational modifications (glucosylations and phosphorylations), cysteines putatively involved in 3D structure, Cu-binding histidines, and consensus sequences for SODs (figure not shown). The deletion observed in three of the sequences was positioned between the aa 85-92. All experimentally obtained sequences of Cu,Zn SOD clustered coincidentally with sequences OePOlee5-1,3,4 and 8 (Figure 2).

3.4 3D-Modeling of the Complete and Deleted Forms of Cu,Zn SODs and Putative Involvement of Modifications in Allergenicity

Sequences OePOlee5-4 and OePOlee5-3 were used as representative forms of the complete and deleted forms of Cu,Zn SOD, respectively. The modeling approaches performed confirmed that the missing part of the protein in the deleted form (8 aa) matched to an external loop which does not form part of the active centre neither binds to the Cu,Zn atoms (Figure 3).

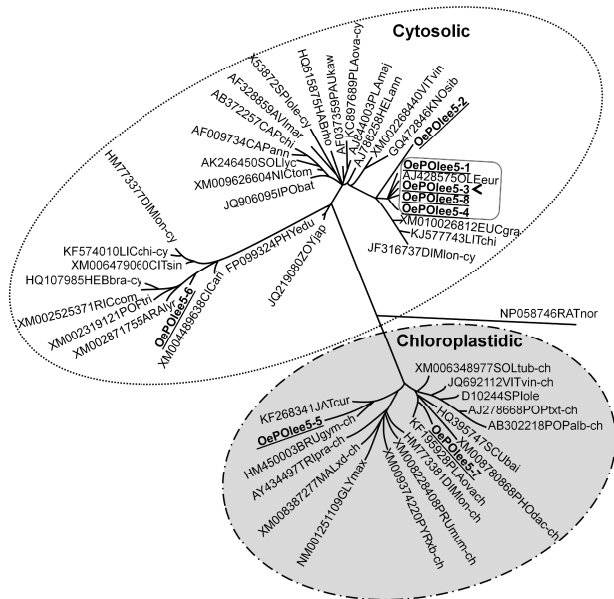


Fig. 2. Phylogenetic tree including olive Cu,Zn SODs sequences obtained either by NGS (underlined) or cloning strategies (which are situated inside the box -not fully listed-), and most relevant heterologous forms from other species (identified by their accession number, shortened name, and -ch or -cy if identified as a chloroplastidic or cytosolic form respectively). Two clusters, corresponding to chloroplastidic and cytosolic sequences are clearly defined. Deleted sequence OePOlee5-3 is pointed out with an arrowhead. The names of the experimentally obtained sequences are not fully displayed. These 72 sequences cluster with OePOlee5-1, 3, 4 and 8 within the rectangle inside the cytosolic cluster. A sequence from rat was included for rooting purposes.

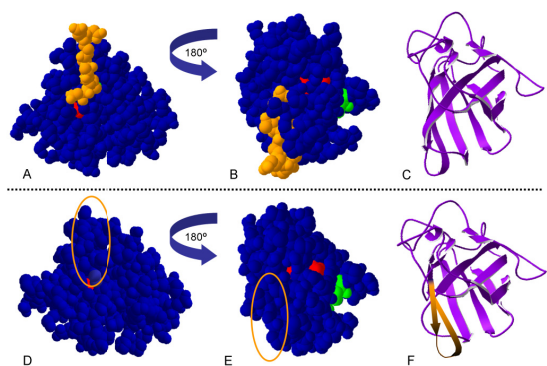


Fig. 3. 3D-modelling of the complete (A-C) and deleted forms (D-F) of the enzyme. The missing sequence is coloured in yellow.

The use of predictive software allowed us to identify the presence of T- and B- epitopes along the two representative forms of SOD mentioned above (OePOlee5-4 and -3). The short sequence (aa 85-92), which is absent from the deleted form, takes part of a defined T-epitope (aa 88-95), a predicted B-linear epitope (aa 87-102), and a conformational T-epitope in the complete sequence. Moreover, the 3-D restructuring occurring in the deleted form when compared to the complete form slightly alters the distribution of other T- and B- epitopes not directly involving the missing loop (results not shown).

4 Discussion

The transcriptomic approach used here greatly helped to define further strategies to identify the variability of the Ole e 5 (Cu,Zn SOD) pollen allergen both within and among olive cultivars, by allowing the identification of the most common variants, which included a deleted form of the enzyme. Such strategies comprised the design of accurate primers for PCR amplification. The presence among pollen transcripts of an even larger number of variants than those expected on the basis of transcriptome analysis was further confirmed by standard cloning procedures, which allowed an experimental validation of the predicted sequences generated by NGS assembly.

Nucleotide polymorphism in the olive pollen allergen Ole e 5 (Cu,Zn SOD) is however relatively lower than that of other olive pollen allergens like Ole e 1 [26] and the considered “highly conserved” allergen Ole e 2 [27] (profilin). Microheterogeneities observed in the sequence cannot be considered artefacts as the result of the misincorporation of nt by the polymerase as they have been detected in Cu,Zn SODs at much higher rate than the described for the enzyme. The observed substitutions do not affect key amino acids for the structure and/or function of the enzyme, like Cys involved in disulphide bridges, His residues involved in Cu-binding, of those motifs described as consensus for Cu,Zn SODs. Therefore, it is likely that the resulting gene products may represent active enzymes. The deleted forms of Cu,Zn SOD observed in both the sequences assembled after NGS as well as after experimental cloning developed in this work are likely not to be greatly affected in their functionality, as neither the reading frame nor the presence of key amino acids were disturbed. However, it is conceivable the presence of slight modifications either in the molecular weight, the isoelectric point of other properties of the protein. Moreover, subtle changes in the activity of the protein, including its kinetics could occur, and should be experimentally tested further.

As regard to the antigenicity/allergenicity of the deleted form of Cu,Zn SOD, predictive tools clearly report a modification of the overall ability of the protein to be recognized as a potential antigen/allergen. T-cell epitopes play a crucial role in immune responses for the induction of cytotoxic T-cell responses and in providing help to B cells for the development of antibody responses, whereas the identification of B-cell epitopes contributes to improve our understanding of structural aspects of

allergens and of the pivotal role they play in the induction of hypersensitivity reactions against small molecules and allergens [28]. The observed differences in the distribution of T- and B-epitopes among the complete and deleted sequences of Cu,Zn SODs may help to develop biotechnological approaches aimed to improve allergy vaccines. These approaches include the generation of hypoallergenic variants including isoforms and folding variants and the engineering of vaccines combining B- or T-cell epitopes. Most recent studies at this regard, resulting in an increased safety profile and reduced side-effects compared with allergen extracts have been recently reviewed [28].

5 Conflict of Interest

The authors confirm that this article content has no conflicts of interest.

Acknowledgments. This work was supported by ERDF-cofunded projects BFU2011-22779 (MINECO), and P2010-AGR-6274, P10-CVI-6075 and P2011-CVI-7487 (Junta de Andalucía). A. Zafra thanks the Agrifood Campus of International Excellence ceiA3 for grant funding. This research was also partially supported by the European Research Program MARIE CURIE (FP7-PEOPLE-2011-IOF), under the grant reference number PEOF-GA-2011-301550 to JCJ-L and JDA. We also thank the I.F.A.P.A. “Venta del Llano” (Mengibar, Jaén, Spain) for providing the plant material.

References

1. Busquets, J., Cour, P., Guerin, B., Michel, F.B.: Busquets et al 1984 Allergy in the Mediterranean area_ I_ Pollen counts and pollinosis (1984)
2. Rodríguez, R., Villalba, M., Monsalve, R.I., Batanero, E.: The spectrum of olive pollen allergens. *Int. Arch. Allergy Immunol.* 125, 185–195 (2001)
3. Jimenez-Lopez, J.C., Rodriguez-Garcia, M.I.: Olive tree genetic background is a major cause of pofilin (Ole e 2 allergen) polymorphism and functional and allergenic variability. *Commun. Agric. Appl. Biol. Sci.* 78, 213–219 (2013)
4. Tejera, M.L., Villalba, M., Batanero, E.: Mechanisms of allergy Identification, isolation, and characterization of Ole e 7, a new allergen of olive tree pollen. *Mech. Allergy* 797–802 (1999)
5. Villalba, M., Rodríguez, R., Batanero, E.: The spectrum of olive pollen allergens. From structures to diagnosis and treatment. *Methods* 66, 44–54 (2014)
6. Batanero, E., Rodríguez, R., Villalba, M.: *Olives and Olive Oil in Health and Disease Prevention*. Elsevier (2010)
7. Boluda, L., Alonso, C., Fernández-Caldas, E.: Purification, characterization, and partial sequencing of two new allergens of *Olea europaea*. *J. Allergy Clin. Immunol.* 101, 210–216 (1998)
8. Butteroni, C., Afferni, C., Barletta, B., Iacovacci, P., Corintia, S., Brunetto, B., Tinghino, R., Ariano, R., Panzani, R.C., Pini, C., Di Felice, G.: Cloning and expression of the olea europaea allergen Ole e 5, the pollen Cu/Zn Superoxide Dismutase (2005)

9. Alché, J.D., Corpas, F.J., Rodríguez-García, M.I., Del Río, L.A.: Identification and immunolocalization of SOD isoforms in olive pollen. *Physiol. Plant* 104, 772–776 (1998)
10. Parida, A.K., Das, A.B.: Salt tolerance and salinity effects on plants: a review. *Ecotoxicol. Environ. Saf.* 60, 324–349 (2005)
11. Alscher, R.G., Erturk, N., Heath, L.S.: Role of superoxide dismutases (SODs) in controlling oxidative stress in plants. *J. Exp. Bot.* 53, 1331–1341 (2002)
12. Corpas, F.J., Fernández-Ocaña, A., Carreras, A., Valderrama, R., Luque, F., Esteban, F.J., Rodríguez-Serrano, M., Chaki, M., Pedrajas, J.R., Sandalio, L.M., del Río, L.A., Barroso, J.B.: The expression of different superoxide dismutase forms is cell-type dependent in olive (*Olea europaea* L.) leaves. *Plant Cell Physiol.* 47, 984–994 (2006)
13. Zafra, A., Traverso, J.A., Corpas, F.J., Alché, J.D.: Peroxisomal localization of CuZn superoxide dismutase in the male reproductive tissues of the olive tree. *Microsc. Microanal.* 18, 2011–2012 (2012)
14. Alché, J.D., Castro, A.J., Jiménez-López, J.C., Morales, S., Zafra, A., Hamman-Khalifa, A.M., Rodríguez-García, M.I.: Differential characteristics of olive pollen from different cultivars: biological and clinical implications. *J. Investig. Allergol. Clin. Immunol.* 17(Suppl 1), 17–23 (2007)
15. Zafra, A., Rodríguez-García, M.I., Alché, J.D.: Cellular localization of ROS and NO in olive reproductive tissues during flower development. *BMC Plant Biol.* (2010)
16. Iba, K.: Acclimative response to temperature stress in higher plants: approaches of gene engineering for temperature tolerance. *Annu. Rev. Plant Biol.* 53, 225–45 (2002)
17. Sambrook, J., Russell, D.: *Molecular Cloning: a laboratory manual*. Cold Spring harbor Laboratory Press, New York (2001)
18. Falgueras, J., Lara, A.J., Fernández-pozo, N., Cantón, F.R., Pérez-trabado, G., Claros, M.G.: SeqTrim: a high-throughput pipeline for pre-processing any type of sequence read. *BMC Bioinformatics* (2010)
19. Gouy, M., Guindon, S., Gascuel, O.: SeaView version 4: A multiplatform graphical user interface for sequence alignment and phylogenetic tree building. *Mol. Biol. Evol.* 27, 221–224 (2010)
20. Abascal, F., Zardoya, R., Posada, D.: ProtTest: selection of best-fit models of protein evolution. *Bioinformatics* 21, 2104–2105 (2005)
21. Guex, N., Peitsch, M.C.: SWISS-MODEL and the Swiss-PdbViewer: An environment for comparative protein modeling (1997)
22. Benkert, P., Biasini, M., Schwede, T.: Toward the estimation of the absolute quality of individual protein structure models. *Bioinformatics* 27, 343–350 (2011)
23. Arnold, K., Bordoli, L., Kopp, J., Schwede, T.: The SWISS-MODEL workspace: a web-based environment for protein structure homology modelling. *Bioinformatics* 22, 195–201 (2006)
24. Schwede, T.: SWISS-MODEL: an automated protein homology-modeling server. *Nucleic Acids Res.* 31, 3381–3385 (2003)
25. Jimenez-Lopez, J.C., Kotchoni, S.O., Hernandez-Soriano, M.C., Gachomo, E.W., Alché, J.D.: Structural functionality, catalytic mechanism modeling and molecular allergenicity of phenylcoumaran benzylic ether reductase, an olive pollen (Ole e 12) allergen. *J. Comput. Aided. Mol. Des.* 27, 873–895 (2013)
26. Hamman-Khalifa, A., Castro, A.J., Jiménez-López, J.C., Rodríguez-García, M.I., Alché, J.D.: Olive cultivar origin is a major cause of polymorphism for Ole e 1 pollen allergen. *BMC Plant Biol.* 8, 10 (2008)

27. Jimenez-Lopez, J.C., Morales, S., Castro, A.J., Volkmann, D., Rodríguez-García, M.I., Alché, J.D.D.: Characterization of profilin polymorphism in pollen with a focus on multifunctionality. *PLoS One* 7, e30878 (2012)
28. Palomares, O., Cramer, R., Rhyner, C.: The contribution of biotechnology toward progress in diagnosis, management, and treatment of allergic diseases. *Allergy* 69, 1588–1601 (2014)