Systematic and Phylogenetic Analysis of the Ole e 1 Pollen Protein Family Members in Plants

José Carlos Jiménez-López,
María Isabel Rodríguez-García and Juan de Dios Alché
Department of Biochemistry, Cell and Molecular Biology of Plants
Estación Experimental del Zaidín, CSIC, Granada, Spain

1. Introduction

Pollen allergens are specific substances able to cause IgE-mediated hypersensitivity (allergy) after contact with the immune system [D’Amato et al. 1998]. To date, about 50 plant species have been registered in the official allergen list of the International Union of Immunological Societies (IUIS) Allergen Nomenclature Subcommittee http://www.allergen.org as capable of inducing pollen allergy in atopic individuals [Mothes et al. 2004]. These plants are usually grouped as (1) trees (members of the orders: Fagales, Pinales, Rosales, Arecales, Scrophulariales, Junglandales, Salicales, and Myrtales), (2) grasses (members of the families: Bambusioideae, Arundinoideae, Chloridoideae, Panicoideae, and Poideae), and (3) weeds (components of families Asteraceae, Chenopodiaceae and Urticaceae) [Hauser et al. 2010].

Allergens are proteins with a broad range of molecular weights (~5 to 50 kDa), which exhibit different features of solubility and stability. More than 10 groups of pollen allergens have been reported. Among all groups of pollen allergens, Pollen Ole e 1 (Ole) domain-containing proteins are the major allergens, included like-members of the “pollen proteins of the Ole e 1 family” (Accession number: PF01190) within the Pfam protein families database [Finn et al. 2010]. Ole e 1 was the first allergen purified from Olea europaea L. [Lauzurica et al. 1998] and named as such according to the IUIS nomenclature [King et al. 1994]. This protein is considered the major olive pollen allergen on the basis of its high prevalence among atopic patients and the high proportion it represents within the total pollen protein content, in comparison with other olive pollen allergens. These include at present another 10 allergens already identified and classified like Ole e 2 to Ole e 11 [Rodríguez et al. 2002, Barral et al. 2004, Salamanca et al. 2010]. Ole e 1 consists of a single polypeptide chain of 145 amino acid residues with a MW of 18–22 kDa, displaying acidic pI and different forms of N-glycosylation [Villalba et al. 1990, Batanero et al. 1994]. Heterologous proteins with a relevant homology have been described in other members of the Oleaceae family, such as fraxinus, lilac, jasmine and privet. The polypeptides encoded by the LAT52 gene from tomato and the Zmc13 gene from maize pollens also exhibit a high similarity to Ole e 1 [Twell et al. 1989, Hanson et al. 1989]. These plant pollen proteins are structurally related but their biological function is not yet known; though they have been suggested to be...
involved in important events of pollen physiology, such as hydration, germination and/or pollen tube growth, and other reproductive functions [Alché et al. 1999, 2004, Tang et al. 2000, Stratford et al. 2001].

Structurally, the Ole domain contains six conserved cysteines which may be involved in disulfide bonds, since no free sulphhydryl groups have been detected in the native protein [Villalba et al. 1993]. Olive Ole e 1 exhibits a high degree of microheterogeneity, mainly concentrated in the third of the molecule closer to the N-terminus. The Ole e I (Ole) domain defining the pollen proteins Ole e I family signature or consensus pattern sequences PS00925 [Sigrist et al. 2010], is characterized by the amino acid sequence [EQT]-G-x-V-Y-C-D-[TNP]-C-R, where “x” could be any residue.

There is a high diversity of proteins sharing the Ole domain among plant species. To date, eleven Ole domain-containing genes have been isolated and characterized from olive pollens [Rodríguez et al. 2002]. Ole-containing proteins include proline-rich proteins, proteins encoding extensin-like domains, phosphoglycerate mutase, tyrosine-rich hydroxyproline-rich glycoprotein, and hydroxyproline-rich glycoprotein. These Ole-containing proteins can exhibit: (1) the pollen Ole signature exclusively, e.g. the ALL1_OLEEU P19963 protein from *Olea europaea* L., (2) both the pollen Ole signature and the replication factor A protein 3 motive pattern (PF08661), e.g. the O49527 pollen-specific protein-like from *Arabidopsis thaliana* (842 residues), (3) both the pollen Ole domain and the phosphoglycerate mutase (PGAM) motif, e.g. the Q9SGZ6 protein from *Arabidopsis thaliana*., and finally (4) both the pollen Ole signature and the reverse transcriptase 2 (RVT2) motif, e.g. the A5AJL0 protein from *Vitis vinifera*.

Several efforts have been made to develop an understandable and reliable systematic classification of the diverse and increasing number of different allergen protein structures. As mentioned above, the classification system widely established for proteins that cause IgE-mediated atopic allergies in humans (allergens) was defined by Chapman et al. (2007). This system uses the first three letters of the genus; a space; the first letter of the species name; a space and an Arabic number. Despite this classification system, protein databases are full of allergen proteins lacking this systematic and comprehensive nomenclature. In other cases, many of the proteins described here have not been described as allergens, or their naming makes no reference to the Ole e 1 family that facilitates their identification. Otherwise, naming in databases is frequently given randomly, on the basis of chromosome location, addressing structural features and functional characterizations or simply using the name of the entire family. In this study, we used a combination of functional genomics and computational biology to name and classify the entire Ole e 1 family, as well as to characterize structurally and functionally the proteins of this superfamily. Our data indicate that the Ole e 1 protein family consists of at least 109 divergent families, which will likely expand as more genomic studies are undertaken, and fully sequenced plant genomes become available.

2. Material and methods

2.1 Database search for Ole e 1 family genes

Sequences of Ole e 1 and Ole e 1-like genes were retrieved from the US National Center for Biotechnology Information (NCBI, http://www.ncbi.nlm.nih.gov/), the Uniprot database (http://www.uniprot.org/), and the non-redundant expressed sequence tag (EST)
databases using BLASTX, BLASTN and BLAST (low complexity filter, Blosum62 substitution matrix) [Altschul et al. 1997]. Searches were conducted using previously characterized *Olea europaea* L. Ole e 1 (GenBank Accession number P19963), *Solanum lycopersicum* LAT52 (GenBank Accession number P13447), *Zea mays* Zmc13 (GenBank Accession number B6T1A9), *Arabidopsis thaliana* pollen-specific protein-like (GenBank Accession number O49527), *Arabidopsis thaliana* PGAM containing domain protein (GenBank Accession number Q9SGZ6), and *Vitis vinifera* RVT2 containing domain protein (GenBank Accession number A5AJL0). Full-length amino acid sequences for Ole e 1 proteins were compiled and aligned using ClustalW [Thompson et al. 1994]. Genetic distances between pairs of amino acid sequences were calculated with Bioedit V7.0.5.3 [Hall 1999]. Consensus protein sequences were derived from these original alignment, and further analyzed for the presence of putative functional motifs using the PROSITE database [Sigrist et al. 2010], of biologically meaningful motif descriptors derived from multiple alignments and the ScanProsite program [de Castro et al. 2006], from the Expert Protein Analysis System (ExPASy) proteomics server of the Swiss Institute of Bioinformatics [Gasteiger et al. 2003]. Finally, the consensus protein sequences were submitted to BLASTP analysis to identify homologous proteins from other plant species.

2.2 Revised/unified nomenclature
In order to provide a revised and unified nomenclature for Ole e 1-like gene superfamily, we developed a sequence-based similarity approach to classify all the retrieved sequences using a previously developed gene nomenclature model [Kotchoni et al. 2010]. For this new nomenclature, Ole e 1 protein sequences that are more than 40% identical to previously identified Ole e 1 sequences compose a family, and sequences more than 60% identical within a family, compose a gene subfamily. Protein sequences that are less than 40% identical would describe a new Ole e 1 gene family. Taking olive protein Ole e 1_57A9 (previous name Ole e 1, major olive pollen allergen) as an example for the revised nomenclature (Table 1), Ole e 1 indicates the root; the digits (57) indicates a family and the first letter (A) a subfamily, while the final number (9) identifies an individual gene within a subfamily. The revised nomenclature is therefore composed of an assigned gene symbol (Ole e 1) (abbreviated gene name) for the whole gene superfamily. The gene symbol must be (i) unique and representative of the gene superfamily; (ii) contain only Latin letters and/or Arabic numerals, (iii) not contain punctuation, and (iv) without any reference to species. These newly developed criteria have been applied to database curators to generate the unified Ole e 1 gene families/classes regardless of the source of the cloned gene(s).

2.3 Sequence alignments and phylogenetic analyses
The retrieved Ole e 1 protein families were used to generate a phylogenetic tree using ClustalW [Thompson et al. 1994]. The alignment was created using the Gonnet protein weight matrix, multiple alignment gap opening/extension penalties of 10/0.5 and pairwise gap opening/extension penalties of 10/0.1. These alignments were adjusted using Bioedit V7.0.5.3 [Hall 1999]. Portions of sequences that could not be reliably aligned were eliminated. Phylogenetic tree was generated by the neighbourjoining method (NJ), and the branches were tested with 1,000 bootstrap replicates. The three was visualized using Treedyn program [Chevenet et al. 2006].
2.4 Ole e 1 superfamily: Protein modeling and structural characterization

In order to study the structural and conformational variability between the Ole e 1 protein families, selected members of the Ole e 1 superfamily were modelled using SWISS-MODEL server, via the ExPASy web server [Gasteiger et al. 2003]. The initial modelled Ole e 1 structures were subjected to energy minimization with GROMOS96 force field energy [van Gunsteren et al. 1996] implemented in DeepView/Swiss-PDBViewer v3.7 [Guex and Peitsch 1997] to improve the van der Waals contacts and to correct the stereochemistry of the improved models. The quality of the models was assessed by checking the protein stereology with PROCHECK [Laskowski et al. 1993] and the protein energy with ANOLEA [Melo et al. 1997, 1998]. Ramachandran plot statistics for the models were calculated to show the number of protein residues in the favoured regions.

3. Results

3.1 The Ole e 1 protein families: Revised and unified nomenclature

In order to provide a revised/international consensus and unified nomenclature for the Ole e 1 gene superfamily, we first retrieved all the Ole e 1 and Ole e 1-like gene sequences using PS00925 as the major molecular consensus defining the entire superfamily of Ole e 1 proteins. We next verified all annotated plant Ole e 1 open reading frames (ORFs) using Ole e 1 sequence domains. A complementary and comparative study was developed by using Uniprot database to validate the molecular function and previous denomination of each Ole e 1 protein. Our searches resulted in the identification of 571 sequences encoding Ole e 1 and Ole e 1 like proteins from a wide variety of plant species, with the diagnostic motif PS00925 (Table 1). According to the established criteria (see Material and Methods), these sequences integrated 109 Ole e 1 gene families which have been attributed to different functional categories including extensins and extensin-like proteins, proline-rich proteins, hydroxyproline-rich glycoproteins, tyrosine-rich/hydroxyproline-rich glycoproteins, hydrolases, phosphoglycerate mutases, arabinogalactan proteins, etc. (Table 1).

Among the sequences retrieved, Ole e 1_48 is the most extensive family with 63 gene members encoding for different pollen-specific protein C13 homologues, followed by Ole e 1_57 family with 42 gene homologues encoding Ole e 1 (the olive major pollen allergen), Ole e 1_16 with 26 gene members encoding proline-rich proteins, and Ole e 1_52 with 22 members encoding LAT52 homologues (Table 1). The number of Ole e 1 genes greatly varied from one plant species to another. The genus Oryza included the highest number of Ole e 1 genes (143), followed by Arabidopsis with 95 genes (Table 1). At present, more than half of the catalogued Ole e 1 families encoded a single Ole e 1/Ole e 1-like gene, which was in most cases “uncharacterized” (Table 1).

The total number of genes in the Ole e 1 superfamily is expected to increase steadily with time, mainly due to the genomic sequencing of additional species like Olea europaea L. (http://www.gene.es.org/11_proyectos/PROYECTOS.CFM?pg=0106&m=1). Regardless of the plethora of Ole e 1 genes yet to be identified/characterized, their classification and relationship to the entire extended Ole e 1 gene superfamily will be easy owing to this nomenclature building block that catalogues newly identified/characterized Ole e 1 gene products only on the basis of sequence similarity to previously characterized Ole e 1 gene products.
**Table 1.** The Ole e 1 protein superfamily: new and unified nomenclature. ARATH: Arabidopsis thaliana; ARALY: Arabidopsis lyrata; BETPN: Betula pendula; BRAOL: Brassica oleracea; BRARP: Brassica rapa; CAPAN: Capsicum annuum; CARAS: Cardaminopsis arenosa; CHE1: Chenopodium album; CROSA: Crocus sativus; DAUCA: Daucus carota; EUPPU: Euphorbia pulcherrima; FRAEX: Fraxinus excelsior; GOSBA: Gossypium barbadense; GOSHE: Gossypium herbaceum; GOSHI: Gossypium hirsutum; GOSKI: Gossypioides kirkii; HYAOR: Hyacinthus orientalis; LigVu: Ligustrum vulgare; LILLO: Lilium longiflorum; LOLPE: Lolium perenne; MAIZE: Zea mays; MEDTR: Medicago truncatula; NICAL: Nicotiana alata; NICGL: Nicotiana glauca; NicLa: Vitis pseudoreticulata; OleEu: Olea europaea; ORYSI: Oryza sativa; PETCR: Petroselinum crispum; PETHY: Petunia hybrida; PHAVU: Phaseolus vulgaris; PHEPR: Phleum pratense; PHYPA: Physcomitrella patens; PICS: Picea sitchensis; PLALA: Platanus lanceolata; POPTR: Populus trichocarpa; RICCO: Ricinus communis; SALTH: Salsola kali; SAMNI: Sambucus nigra; SELML: Selaginella moellendorffii; SOLL: Solanum lycoctonum; SORBI: Sorgum bicolor; SOYB: Glycine max; TOBAC: Nicotiana tabacum; TRISU: Trifolium subterraneum; VITVI: Vitis vinifera; 9ROSI: Cleome spinosa; (-): uncharacterized.
Table 1. (continued). The Ole e 1 protein superfamily: new and unified nomenclature.

| 14  | Csa e 3  | ABA00911.1 | AB0129.2 | AEN7.5 | AEN7.5 | AEN7.5 |
|-----|----------|-------------|----------|--------|--------|--------|
| 15  | Csa e 3  | ABA00905.1 | AB0129.2 | AEN7.5 | AEN7.5 | AEN7.5 |
| 16  | Csa e 3  | ABA00900.1 | AB0129.2 | AEN7.5 | AEN7.5 | AEN7.5 |
| 17  | Csa e 3  | ABA00895.1 | AB0129.2 | AEN7.5 | AEN7.5 | AEN7.5 |
| 18  | Csa e 3  | ABA00890.1 | AB0129.2 | AEN7.5 | AEN7.5 | AEN7.5 |
| 19  | Csa e 3  | ABA00885.1 | AB0129.2 | AEN7.5 | AEN7.5 | AEN7.5 |
| 20  | Csa e 3  | ABA00880.1 | AB0129.2 | AEN7.5 | AEN7.5 | AEN7.5 |
| 21  | Csa e 3  | ABA00875.1 | AB0129.2 | AEN7.5 | AEN7.5 | AEN7.5 |
| 22  | Csa e 3  | ABA00870.1 | AB0129.2 | AEN7.5 | AEN7.5 | AEN7.5 |
| 23  | Csa e 3  | ABA00865.1 | AB0129.2 | AEN7.5 | AEN7.5 | AEN7.5 |
| 24  | Csa e 3  | ABA00860.1 | AB0129.2 | AEN7.5 | AEN7.5 | AEN7.5 |
| 25  | Csa e 3  | ABA00855.1 | AB0129.2 | AEN7.5 | AEN7.5 | AEN7.5 |
| 26  | Csa e 3  | ABA00850.1 | AB0129.2 | AEN7.5 | AEN7.5 | AEN7.5 |
| 27  | Csa e 3  | ABA00845.1 | AB0129.2 | AEN7.5 | AEN7.5 | AEN7.5 |
| 28  | Csa e 3  | ABA00840.1 | AB0129.2 | AEN7.5 | AEN7.5 | AEN7.5 |
| 29  | Csa e 3  | ABA00835.1 | AB0129.2 | AEN7.5 | AEN7.5 | AEN7.5 |
| 30  | Csa e 3  | ABA00830.1 | AB0129.2 | AEN7.5 | AEN7.5 | AEN7.5 |
| 31  | Csa e 3  | ABA00825.1 | AB0129.2 | AEN7.5 | AEN7.5 | AEN7.5 |
| 32  | Csa e 3  | ABA00820.1 | AB0129.2 | AEN7.5 | AEN7.5 | AEN7.5 |
| 33  | Csa e 3  | ABA00815.1 | AB0129.2 | AEN7.5 | AEN7.5 | AEN7.5 |
| 34  | Csa e 3  | ABA00810.1 | AB0129.2 | AEN7.5 | AEN7.5 | AEN7.5 |
| 35  | Csa e 3  | ABA00805.1 | AB0129.2 | AEN7.5 | AEN7.5 | AEN7.5 |
| 36  | Csa e 3  | ABA00800.1 | AB0129.2 | AEN7.5 | AEN7.5 | AEN7.5 |
| 37  | Csa e 3  | ABA00795.1 | AB0129.2 | AEN7.5 | AEN7.5 | AEN7.5 |
| 38  | Csa e 3  | ABA00790.1 | AB0129.2 | AEN7.5 | AEN7.5 | AEN7.5 |
| 39  | Csa e 3  | ABA00785.1 | AB0129.2 | AEN7.5 | AEN7.5 | AEN7.5 |
| 40  | Csa e 3  | ABA00780.1 | AB0129.2 | AEN7.5 | AEN7.5 | AEN7.5 |
| 41  | Csa e 3  | ABA00775.1 | AB0129.2 | AEN7.5 | AEN7.5 | AEN7.5 |
| 42  | Csa e 3  | ABA00770.1 | AB0129.2 | AEN7.5 | AEN7.5 | AEN7.5 |
| 43  | Csa e 3  | ABA00765.1 | AB0129.2 | AEN7.5 | AEN7.5 | AEN7.5 |
| 44  | Csa e 3  | ABA00760.1 | AB0129.2 | AEN7.5 | AEN7.5 | AEN7.5 |
| 45  | Csa e 3  | ABA00755.1 | AB0129.2 | AEN7.5 | AEN7.5 | AEN7.5 |
| 46  | Csa e 3  | ABA00750.1 | AB0129.2 | AEN7.5 | AEN7.5 | AEN7.5 |
| 47  | Csa e 3  | ABA00745.1 | AB0129.2 | AEN7.5 | AEN7.5 | AEN7.5 |
| 48  | Csa e 3  | ABA00740.1 | AB0129.2 | AEN7.5 | AEN7.5 | AEN7.5 |
| 49  | Csa e 3  | ABA00735.1 | AB0129.2 | AEN7.5 | AEN7.5 | AEN7.5 |
| 50  | Csa e 3  | ABA00730.1 | AB0129.2 | AEN7.5 | AEN7.5 | AEN7.5 |
| 51  | Csa e 3  | ABA00725.1 | AB0129.2 | AEN7.5 | AEN7.5 | AEN7.5 |
| 52  | Csa e 3  | ABA00720.1 | AB0129.2 | AEN7.5 | AEN7.5 | AEN7.5 |
| 53  | Csa e 3  | ABA00715.1 | AB0129.2 | AEN7.5 | AEN7.5 | AEN7.5 |
| 54  | Csa e 3  | ABA00710.1 | AB0129.2 | AEN7.5 | AEN7.5 | AEN7.5 |
| 55  | Csa e 3  | ABA00705.1 | AB0129.2 | AEN7.5 | AEN7.5 | AEN7.5 |
| 56  | Csa e 3  | ABA00700.1 | AB0129.2 | AEN7.5 | AEN7.5 | AEN7.5 |
| 57  | Csa e 3  | ABA00695.1 | AB0129.2 | AEN7.5 | AEN7.5 | AEN7.5 |
| 58  | Csa e 3  | ABA00690.1 | AB0129.2 | AEN7.5 | AEN7.5 | AEN7.5 |
| 59  | Csa e 3  | ABA00685.1 | AB0129.2 | AEN7.5 | AEN7.5 | AEN7.5 |
| 60  | Csa e 3  | ABA00680.1 | AB0129.2 | AEN7.5 | AEN7.5 | AEN7.5 |
| 61  | Csa e 3  | ABA00675.1 | AB0129.2 | AEN7.5 | AEN7.5 | AEN7.5 |
| 62  | Csa e 3  | ABA00670.1 | AB0129.2 | AEN7.5 | AEN7.5 | AEN7.5 |
| 63  | Csa e 3  | ABA00665.1 | AB0129.2 | AEN7.5 | AEN7.5 | AEN7.5 |
| 64  | Csa e 3  | ABA00660.1 | AB0129.2 | AEN7.5 | AEN7.5 | AEN7.5 |
| 65  | Csa e 3  | ABA00655.1 | AB0129.2 | AEN7.5 | AEN7.5 | AEN7.5 |
| 66  | Csa e 3  | ABA00650.1 | AB0129.2 | AEN7.5 | AEN7.5 | AEN7.5 |
| 67  | Csa e 3  | ABA00645.1 | AB0129.2 | AEN7.5 | AEN7.5 | AEN7.5 |

www.intechopen.com
Table 1. (continued). The Ole e 1 protein superfamily: new and unified nomenclature.

| Accession | Description | New Name |
|-----------|-------------|----------|
| A0A01615720 | Ole e 1 allergen | Ole e 1 ( unified) |
| A0A01615730 | Ole e 1 allergen | Ole e 1 ( unified) |
| A0A01615740 | Ole e 1 allergen | Ole e 1 ( unified) |
| A0A01615750 | Ole e 1 allergen | Ole e 1 ( unified) |
| A0A01615760 | Ole e 1 allergen | Ole e 1 ( unified) |
| A0A01615770 | Ole e 1 allergen | Ole e 1 ( unified) |
| A0A01615780 | Ole e 1 allergen | Ole e 1 ( unified) |

www.intechopen.com
Table 1. (continued). The Ole e 1 protein superfamily: new and unified nomenclature.

| Gene symbol | Name | Species | Accession number | Description |
|-------------|------|---------|------------------|-------------|
| Os109| Phalloidin-binding protein| Oryza sativa| Q4B335| NIGAL |
| Os101| Phototactic attractant protein| Oryza sativa| Q8I217| NFRL |
| Os102| Phototactic attractant protein| Oryza sativa| Q8I217| NFRL |
| Os103| Phototactic attractant protein| Oryza sativa| Q8I217| NFRL |
| Os104| Phototactic attractant protein| Oryza sativa| Q8I217| NFRL |
| Os105| Phototactic attractant protein| Oryza sativa| Q8I217| NFRL |
| Os106| Phototactic attractant protein| Oryza sativa| Q8I217| NFRL |
| Os107| Phototactic attractant protein| Oryza sativa| Q8I217| NFRL |
| Os108| Phototactic attractant protein| Oryza sativa| Q8I217| NFRL |
| Os109| Phototactic attractant protein| Oryza sativa| Q8I217| NFRL |
| Os110| Phototactic attractant protein| Oryza sativa| Q8I217| NFRL |
| Os111| Phototactic attractant protein| Oryza sativa| Q8I217| NFRL |
| Os112| Phototactic attractant protein| Oryza sativa| Q8I217| NFRL |
| Os113| Phototactic attractant protein| Oryza sativa| Q8I217| NFRL |
| Os114| Phototactic attractant protein| Oryza sativa| Q8I217| NFRL |
| Os115| Phototactic attractant protein| Oryza sativa| Q8I217| NFRL |
| Os116| Phototactic attractant protein| Oryza sativa| Q8I217| NFRL |
| Os117| Phototactic attractant protein| Oryza sativa| Q8I217| NFRL |
| Os118| Phototactic attractant protein| Oryza sativa| Q8I217| NFRL |
| Os119| Phototactic attractant protein| Oryza sativa| Q8I217| NFRL |
| Os120| Phototactic attractant protein| Oryza sativa| Q8I217| NFRL |
| Os121| Phototactic attractant protein| Oryza sativa| Q8I217| NFRL |
| Os122| Phototactic attractant protein| Oryza sativa| Q8I217| NFRL |
| Os123| Phototactic attractant protein| Oryza sativa| Q8I217| NFRL |

*Note: The table continues with additional entries.*
3.2 Phylogenetic analysis of the extended Ole e 1 protein families
A member of each retrieved full-length Ole e 1 sequences family was aligned to determine phylogenetic relationships within the Ole e 1 extended family. A phylogenetic tree of the Ole e 1 extended sequences is depicted in Figure 1.

Fig. 1. Phylogenetic analysis of plant Ole e 1 proteins. Neighbour-Joining (NJ) method was used to perform a phylogenetic analysis of Ole e 1 proteins from 109 families. One representative sequence of each family was used, based in its higher consensus ability. Plant species analyzed included *Arabidopsis*, poplar, rice, spikemoss, tobacco, maize, potato, grape, *Sorghum*, kidney bean, barrel medic, *Pinus*, poinsettia, perennial ryegrass, soybean, white birch, ash, *Platanus*, *Physcomitrella*, cotton, subterranean clover, Persian tobacco and castor bean.

The phylogenetic tree shows that the 109 Ole e 1 extended families, although highly divergent, are split into two clades. The smaller clade was integrated by a few species like *Selaginella moellendorffii, Arabidopsis* and maize among others. The second clade included the majority of the Ole e 1 family proteins, clustering together almost all the biological functions (Figure 1). Numerous branches aroused from this clade.

3.3 Ole e 1 protein superfamilies: Structural and conformational variability
The crystallographic structural coordinates of relatively few proteins of the Ole e 1 family have been deposited in the Protein Database (PDB) up to date. To our knowledge, detailed comparative studies of the structural and conformational features of members of the Ole e 1
extended protein families have not been performed in higher plants. Using computational modelling analysis, we have determined and modelled the molecular-structural features of selected members of the Ole e 1 extended families. A first overview of the generated models (Figure 2) indicated a relatively high level of similitude.

![Figure 2](https://www.intechopen.com)

Fig. 2. Three-dimensional structure analysis of selected members of Ole e 1 family proteins. The model proteins are depicted as cartoon diagrams. The secondary elements of the crystallographic structures are rainbow coloured, with N-terminus in blue, and C-terminus in red.

However, a more detailed analysis allowed identifying certain differences in the generated models, particularly consisting in 2D structural features. These differences can be distinguished even between very close proteins like P19963, AF532754 and AF532760 (Ole e 1_57A9, Ole e 1_57A25 and Ole e 1_57A23 with the new nomenclature), corresponding to the olive pollen major allergen cloned from different varietal sources or even to different clones of the same cultivar (Figure 2). The differences become higher when models of the same protein obtained from different plant species are compared. This is the case of P13447 and B9SBK9 (Ole e 1_52L1 and Ole e 1_52J1), which correspond to the LAT52 gene product in tomato and *Ricinus communis*, respectively (Figure 2). Divergences are even more obvious between the models indicated above and that of a P33050 (Ole e 1_48H6), a different member of the Ole e 1 superfamily corresponding to a pollen protein from maize (C13 protein) (Figure 2).
4. Discussion

Research as regard to the proteins of the Ole e 1 family has been carried out steadily since its definition. At present, many genes from the allergen Ole e 1 family of proteins have been characterized, and data are available concerning the sequence, structure, expression and biological function (e.g. extensin-like proteins constituting part of the cell wall). However, and as depicted in this chapter, the precise identification of more than half members of this family remains uncompleted. Up to now, Ole e 1 and Ole e 1-like genes are deposited into the databases, many of them with repetitive or arbitrary naming system by authors. This nomenclature includes a variety of generic names, such as Ole e 1 major olive pollen allergen, putative Ole e 1-like protein, anther-specific Ole e 1-like protein, and others depending of the protein location in the chromosome, e.g. At3g26960, Os09g0508200, or simply giving a random name e.g. P1 clone: MOJ10. For those members of the Ole e 1 family which have been recognized like allergens, a more sustainable and precise nomenclature has been built, by following the recommendations of the International Union of Immunological Societies (IUIS) (http://www.allergen.org/). However, these allergenic proteins only represent a part of the members of the Ole e 1 family, and this nomenclature still does not display the relationships among these proteins. In several cases, it is still common for researchers to use different names for the same allergen. Allergen biochemistry is now entering a new time of structural biology and proteomics that will require sophisticated tools for data processing and bioinformatics, and might require further definition of the nomenclature. Increasingly, the wealth of structural information is enabling the biologic function of allergens to be established and the assignment of allergen function to diverse protein families. Therefore, the arbitrary nomenclature currently in use is not sustainable for adequate comparative mega-functional genomics studies, especially as the number of Ole e 1 genes has increased steadily and will continue with this upward trend with the completion of the sequencing projects corresponding to more plant genomes.

The implementation of modifications in the nomenclature as proposed here may assist further developments of allergy understanding and new clinical approaches. As an example, nomenclature and structural biology have been proposed to play a crucial role in defining allergens for research studies and for the development of new clinical products [Chapman et al. 2007]. Sequence comparisons and assignments to protein families provide a molecular basis for clinical cross-reactions between food, pollen, and latex allergens that give rise to oral allergy syndromes [Wagner et al. 2002, Scheiner et al. 2004, van Ree 2004]. For food and pollen allergens, intrinsic protein structure probably plays an important role in determining allergenicity by conferring, for example, heat stability or resistance to digestion in the digestive tract, e.g. storage proteins from seed/nuts or legumes [Orruño and Morgan 2011]. Interestingly, analysis of databases, e.g. pFAM shows that there are currently more than 120 molecular architectures that are responsible for eliciting IgE responses. It will be important to link nomenclature with classification of allergens into protein families and subfamilies to provide complete definition of allergens and their structure-functional relationships as part of a comprehensive bioinformatics database. The practical consequences of this approach are seen most clearly with genetically modified foods, in which sequence comparisons can be used for safety assessment of genetically modified organisms [Goodman and Tetteh 2011].

The success of our new and unified nomenclature lies in its simplicity, with genetic basis and structural-functional characterizations of the proteins, regardless of the species origin,
with the possibility to further nomenclature expansion, to include as-yet-unidentified protein allergens from different sources or species: mites, insects, pollens, molds and foods. It might be also possible to include in the system engineered protein molecules, such as hypoallergens, or others being described as non-protein allergens. Allergens entered into the nomenclature could be used to develop allergen-specific diagnostics and to formulate recombinant allergen vaccines that will benefit patients [Chapman et al. 2000, Ferreira et al. 2004, Jutel et al. 2005, Sastre 2010].

The proposed system may also assist to clarify the importance of allergen polymorphism. Allergens often display numerous variants. These are proteins with typically greater than 90% sequence identity, but with enough differences in their amino acid sequences to make worth individual structural and or functional characterization and identification. This polymorphism has been deeply analyzed in mites, as their allergens present an extensive number of isoforms: 23 for Der p 1 and 13 for Der p 2 [Smith et al. 2001, Smith et al. 2001]. Furthermore, these polymorphisms might affect T-cell responses or alter antibody-binding sites. These differences can be structurally characterized to distinguish isoforms in a well-defined nomenclature system, by mean of structural-functional differentiation, helping to design allergen formulations for immunotherapy [Jutel et al. 2005, Piboonpocanun et al. 2006]. In the case of pollen allergens, Ole e 1 from olive pollen is a clear example of extreme polymorphism, both in its peptide and in its carbohydrate moieties, as demonstrated by peptide mapping and N-glycopeptide analysis [Castro et al. 2010]. Olive cultivar origin is a major cause of polymorphism for Ole e 1 pollen allergen [Hamman-Khalifa et al. 2008, Castro et al. 2010]. The olive tree has an extremely wide germplasm, with over 1200 varieties cultivated over the world [Bartolini et al. 1994]. Therefore, the number of Ole e 1 isoforms yet to be characterized in olive pollen is expected to be enormous. A similar situation is also likely to occur in many other plant species.

Overall, our developed unified nomenclature system is helpful in a quick functional prediction of any newly cloned Ole e 1 gene(s), because from the nomenclature point of view, the newly sequenced gene(s) will always be characterized/named with sequence similarity with previously characterized Ole e 1 genes/proteins, as well as a protein structure-functional characterization and comparison. The changes that have been introduced reflect into which extended family or subfamily a certain Ole e 1 protein belongs. Accordingly, the new nomenclature will have no significant impact on already published data with old/arbitrary naming system. However, we urge scientists working on Ole e 1’s to adopt this new and easy nomenclature system. In this regard, we have made an effort to preserve the user friendly linkage between the old and the new designations, which we hope will help researchers to adapt the new names. As the revised nomenclature should facilitate communication and understanding within the community interested in Ole e 1 allergen proteins, we advocate that this new naming system be used in all future studies.

The classification model used here has been developed under the basis of a previously designed gene nomenclature model for male fertility restorer (RF) proteins in higher plants [Kotchoni et al. 2010]. The increasing numbers of RF genes described in the literature represented an ongoing challenge in their clear identification and logical classification which was solved using the proposed nomenclature. Undoubtedly, similar approaches could be applied to numerous protein families involving relevant levels of nomenclature heterogeneity, many of them registered in specialized databases like pFam. In the case of allergens, other numerous protein families like profilins (Ole e 2 in the case of olive pollen).
prolamins, cupins, Bet v 1-related proteins etc., which are currently included in the AllFam database [Radauer et al. 2008] (http://www.meduniwien.ac.at/allergens/allfam/) could benefit of the use of similar approaches.

5. Conclusion

We propose for first time a unified naming system for Ole e 1-like genes and pseudogenes across all plant species, which accommodates the numerous sequences already deposited in several databases, offering the needed flexibility to incorporate additional Ole e 1-like proteins as they become available. Additionally, we provide an analysis of the phylogenetic relationships displayed by the members of the Ole e 1-like family and use computational protein modelling to determine structural features of selected members of this family. These data are of particular relevance for the understanding of their biological activity and allergenic cross-reactivity.

6. Acknowledgment

Support of the Spanish Ministry of Science and Innovation (ERDF-cofinanced project BFU2008-00629) and Andalusian Regional Government (ERDF-cofinanced Proyectos de Excelencia CVI5767 and AGR6274) is gratefully acknowledged.

7. References

Alché, J.D.; Castro, A.J.; Olmedilla, A.; Fernández, M.C.; Rodríguez, R.; Villalba, M. and Rodríguez-García, M.I. (1999). The major olive pollen allergen (Ole e 1) shows both gametophytic and sporophytic expression during anther development, and its synthesis and storage takes place in the RER. *Journal of Cell Science*, Vol.112, pp.2501-2509

Alché, J.D.; Mrani-Alaoui, M.; Castro, A.J. and Rodríguez-García, M.I. (2004). Ole e 1, the major allergen from olive (*Olea europaea* L.) pollen, increases its expression and is released to the culture medium during in vitro germination. *Plant Cell Physiology*, Vol.45, pp.1149-1157

Altschul, S.F.; Gish, W.; Miller, W.; Myers, E.W. and Lipman, D.J. (1990). Basic local alignment search tool. *Journal of Molecular Biology*, Vol.215, No.3, pp.403-410

Altschul, S.F.; Madden, T.L.; Schäffer, A.A.; Zhang, J.; Zhang, Z.; Miller, W. and Lipman, D.J. (1997). Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. *Nucleic Acids Research*, Vol.25, No.17, pp.3389-402

Barral, P.; Batanero, E.; Palomares, O.; Quiralte, J.; Villalba, M. and Rodríguez, R. (2004). A major allergen from pollen defines a novel family of plant proteins and shows intra- and interspecies cross-reactivity. *Journal of Immunology*, Vol.172, pp.3644-3651

Bartolini, G.; Prevost, G. and Messeri, C. (1994). Olive tree germplasm: descriptor lists of cultivated varieties in the world. *Acta Horticulturae*, Vol.365, pp.116-118

Batanero, E.; Villalba, M. and Rodríguez, R. (1994). Glycosylation site of the major allergen from olive tree. Allergenic implications of the carbohydrate moiety. *Molecular Immunology*, Vol.31, pp.31-37
Castro, A.J.; Bednarczyk, A.; Schaeffer-Reiss, C.; Rodriguez-Garcia, M.I.; Van Dorsselaer, A.; Alché, J.D. (2010). Screening of Ole e 1 polymorphism among olive cultivars by peptide mapping and N-glycopeptide analysis. *Proteomics*, Vol. 10, No 5, pp.953-962

Chapman, M.D.; Pomés, A.; Breiteneder, H. and Ferreira, F. (2007). Nomenclature and structural biology of allergens. *Journal of Allergy and Clinical Immunology*, Vol.119, No.2, pp.414-420

Chapman, M.D.; Smith, A.M.; Vailes, L.D.; Arruda, K.; Dhanaraj, V. and Pomes, A. (2000). Recombinant allergens for diagnosis and therapy of allergic diseases. *The Journal of Allergy and Clinical Immunology*, Vol.106, pp.409-418

Chevenet, F.; Brun, C.; Banuls, A.L.; Jacq, B. and Christen, R. (2006). TreeDyn: towards dynamic graphics and annotations for analyses of trees. *BMC Bioinformatics*, Vol.7, pp.439

D’Amato, G.; Spieksma, F.T.; Liccardi, G.; Jager, S.; Russo, M.; Kontou-Fili, K.; Nikkels, H.; Wuthrich, B. and Bonini, S. (1998). Pollen-related allergy in Europe. *Allergy*, Vol.53, pp.67-78

de Castro, E.; Sigrist, C.J.A.; Gattiker, A.; Bulliard, V.; Langendijk-Genevaux, P.S.; Gasteiger, E.; Bairoch, A. and Hulo, H. (2006) ScanProsite: detection of PROSITE signature matches and ProRule-associated functional and structural residues in proteins. *Nucleic Acids Research*, Vol.34, pp.362-365

Ferreira, F.; Wallner, M. and Thalhammer, J. (2004). Customized antigens for desensitizing allergic patients. *Advances in Immunology*, Vol.84, pp.79-129

Finn, R.D.; Mistry, J.; Tate, J.; Coggill, P.; Heger, A.; Pollington, J.E.; Gavin, O.L. Gunasekaran, P.; Ceric, G. Forslund, K.; Holm, L.; Sonnhammer, E.L.; Eddy, S.R. and Bateman, A. (2010). The Pfam protein families database. *Nucleic Acids Research*, Database Issue 38, pp.D211-222

Gasteiger, E.; Gattiker, A.; Hoogland, C.; Ivanyi, I.; Appel R.D. and Bairoch A. (2003) ExPASy: the proteomics server for in-depth protein knowledge and analysis. *Nucleic Acids Research*, Vol.31, pp.3784–3788

Goodman, R.E. and Tetteh, A.O. (2011). Suggested Improvements for the Allergenicity Assessment of Genetically Modified Plants Used in Foods. *Current Allergy and Asthma Reports*, doi: 10.1007/s11882-011-0195-6

Guex, N. and Peitsch, M.C. (1997). SWISS-MODEL and the Swiss-PdbViewer: an environment for comparative protein modeling. *Electrophoresis*, Vol.18, No.15, pp.2714–2723

Hall, T.A. (1999). BioEdit: a user-friendly biological sequence alignment editor and analysis program for Windows 95/98/NT. *Nucleic Acids Symposium Series*, Vol.41, pp.95-98

Hamman-Khalifa, A.M.; Castro A.J.; Jimenez-Lopez, J.C.; Rodríguez-Garcia, M.I. and Alché, J.D. (2008). Olive cultivar origin is a major cause of polymorphism for Ole e 1 pollen allergen. *BMC Plant Biology*, Vol.8, 10

Hanson, D.D.; Hamilton, D.S.; Travis, J.L.; Bashe, D.M. and Mascarenhas, J.P. (1998). Characterization of a pollen-specific cDNA clone from Zea mays and its expression. *Plant Cell*, Vol.1, pp.173-179

Hauser, M.; Roulias, A.; Ferreira, F. & Egger, M. (2010). Panallergens and their impact on the allergic patient. *Allergy, Asthma & Clinical Immunology*, Vol.6, pp.1-
Jutel, M.; Jaeger, L.; Suck, R.; Meyer, H.; Fiebig, H. and Cromwell, O. (2005). Allergenspecific immunotherapy with recombinant grass pollen allergens. *The Journal of Allergy and Clinical Immunology*, Vol.116, pp.608-613

Laskowski, R.A.; MacArthur, M.W.; Moss, D.S. and Thornton, J.M. (1993). PROCHECK: A program to check the stereo-chemical quality of protein structures. *Journal of Applied Crystallography*, Vol.26, pp.283–291

Lauzurica, P.; Gurbindo, C.; Maruri, N.; Galocha, B.; Diaz, R.; Gonzalez, J.; Garcia, R. and Lahoz, C. (1988). Olive (*Olea europea*) pollen allergens—I. Immunochemical characterization by immunoblotting, CRIE and immunodetection by a monoclonal antibody. *Molecular Immunology*, Vol.25, pp.329-335

King, T.P.; Hoffman, D.; Lowenstein, H.; Marsh, D.G.; Platts-Mills, T.A. and Thomas, W. (1994). Allergen nomenclature. WHO/IUIS Allergen Nomenclature Subcommittee. *International Archives of Allergy and Immunology*, Vol. 105, pp. 224-233

Kotchoni, S.O.; Jimenez-Lopez, J.C.; Gachomo, W.E. and Seufferheld, M.J. (2010). A new and unified nomenclature for male fertility restorer (RF) proteins in higher plants. *PLoS ONE*, Vol.5, No.12, pp.e15906

Melo, F. and Feytmans, E. (1997). Novel knowledge-based mean force potential at atomic level. *Journal of Molecular Biology*, Vol.267, No.1, pp.207-222

Melo, F. and Feytmans, E. (1998). Assessing protein structures with a non-local atomic interaction energy. *Journal of Molecular Biology*, Vol.277, No.5, pp.1141-1152

Mothes, N.; Horak, F. & Valenta, R. (2004). Transition from a botanical to a molecular classification in tree pollen allergy: implications for diagnosis and therapy. *International Archives of Allergy and Immunology*, Vol.135, pp.357-373

Orruño, E. and Morgan, M.R.A. (2011). Resistance of purified seed storage proteins from sesame (*Sesamum indicum* L.) to proteolytic digestive enzymes. *Food Chemistry*, in press

Piboonpocanun S, Malinual N, Jirapongsananuruk J, Vichyanond P, Thomas WR. (2006). Genetic polymorphisms of major house dust mite allergens. *Clinical & Experimental Allergy*, Vol.36, pp.510-516

Radauer, C.; Bublin, M.; Wagner, S.; Mari, A. and Breiteneder, H. (2008). Allergens are distributed into few protein families and possess a restricted number of biochemical functions. *Journal of Allergy and Clinical Immunology*, Vol.121, pp.847-852

Rodriguez, R.; Villalba, M.; Batanero, E.; González, E.M.; Monsalve, R.I.; Huecas, S.; Tejera, M.L. and Ledesma, A. (2002). Allergenic diversity of the olive pollen. *Allergy*, Vol.57, pp.6-16

Rodriguez, R.; Villalba, M.; Monsalve, R.I.; Batanero, E.; González, E.M.; Monsalve, R.I.; Huecas, S.; Tejera, M.L. and Ledesma, A. (2002). Allergenic diversity of the olive pollen. *Allergy*, Vol.57, pp.6-15

Salamanca, G.; Rodriguez, R. Quiralte, J.; Moreno, C.; Pascual, C.Y.; Barber, D. and Villalba, M. (2010). Pectin methylesterases of pollen tissue, a major allergen in olive tree. *FEBS Journal*, Vol.277, No.13, pp.2729-2739

Sastre, J. (2010). Molecular diagnosis in allergy. *Clinical & Experimental Allergy*, Vol.40, No.10, pp.1442-1460

Scheiner, O.; Aberer, W.; Ebner, C.; Ferreira, F.; Hoffmann-Sommergruber, K.; Hsieh, L.S.; Kraft, D.; Sowka, S.; Vanek-Krebitz, M. and Breiteneder, H. (1997). Cross-racting
allergens in tree pollen and pollen-related food allergy: implications for diagnosis of specific IgE. International Archives of Allergy and Immunology, Vol.113, pp.105-108

Shultz, J.L.; Kurunam, D.; Shopinski, K.; Iqbal, M.J.; Kazi, S.; Zobrist, K.; Bashir, R.; Yaegashi, S.; Lavu, N.; Afzal, A.J.; Yesudas, C.R.; Kassem, M.A.; Wu, C.; Zhang, H.B.; Town, C.D.; Meksem, K. and Lightfoot, D.A. (2006). The Soybean Genome Database (SoyGD): a browser for display of duplicated, polyploid, regions and sequence tagged sites on the integrated physical and genetic maps of Glycine max. Nucleic Acids Research, Vol.34(suppl 1), pp.D758-D765

Sigrist, C.J.A.; Cerutti, L.; de Castro, E.; Langendijk-Genevaux, P.S.; Bulliard, V.; Bairoch, A. and Hulo, N. (2010). PROSITE, a protein domain database for functional characterization and annotation. Nucleic Acids Research, Vol.38 (Database issue), pp.161-166

Smith, A.S.; Benjamin, D.C.; Hozic, N.; Derewenda, U.; Smith, W.A.; Thomas, W.R.; Gafvelin, G.; van Hage-Hamsten, M. and Chapman, M.D. (2001). The molecular basis of antigenic cross-reactivity between the group 2 mite allergens. The Journal of Allergy and Clinical Immunology, Vol.107, pp.977-984

Smith, W.A.; Hales, B.J.; Jarnicki, A.G. and Thomas W.R. (2001). Allergens of wild house dust mites: environmental Der p 1 and Der p 2 sequence polymorphisms. The Journal of Allergy and Clinical Immunology, Vol.107, pp.985-992

Stratford, S.; Barne, W.; Hohorst, D.L.; Sagert, J.G.; Cotter, R.; Golubiewski, A.; Showalter, A.M.; McCormick, S. and Bedinger, P. (2001). A leucine-rich repeat region is conserved in pollen extensin-like (Pex) proteins in monocots and dicots. Plant Molecular Biology, Vol.46, pp.43-56

Tang, B.; Banerjee, B.; Greenberger, P.A.; Fink, J.N.; Kelly, K.J. and Kurup, V.P. (2000). Antibody binding of deletion mutants of Asp f 2, the major Aspergillus fumigatus allergen. Biochemical and Biophysical Research Communications, Vol.270, pp.1128-1135

Thompson, J.D.; Higgins, D.G. and Gibson, T.J. (1994). CLUSTAL W: improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position-specific gap penalties and weight matrix choice. Nucleic Acids Research, Vol.22, pp.4673–4680

Twell, D.; Wing, R.; Yamaguchi, J. and McCormick, S. (1989). Isolation and expression of an anther-specific gene from tomato. Molecular and General Genetics, Vol.217, pp.240-245

van Gunsteren, W.F.; Billeter, S.R.; Eising, A.A.; Hünenberger, P.H.; Krüger, P.; Mark, A.E.; Scott, W.R.P. and Tironi, I.G. (1996). Biomolecular Simulations: The GROMOS96 Manual and User Guide. Zürich, VdF Hochschulverlag ETH Z

van Ree R. (2004). Clinical importance of cross-reactivity in food allergy. Current Opinion in Allergy & Clinical Immunology. Vol.4, pp.235-240

Villalba, M.; Batanero, E.; Lopez-Otin, C.; Sanchez, L.M.; Monsalve, R.I.; Gonzalez de la Pena, M.A.; Lahoz, C. and Rodriguez, R. (1993). The amino acid sequence of Ole e I, the major allergen from olive tree (Olea europaea) pollen. European Journal of Biochemistry, Vol.216, pp.863-869

Villalba, M.; López-Otin, C.; Martín-Orozco, E.; Monsalve, R.I.; Palomino, P.; Lahoz, C. and Rodriguez, R. (1990). Isolation of three allergenic fractions of the major allergen from Olea europaea pollen and N-terminal amino acid sequence. Biochemical and Biophysical Research Communications, Vol.172, pp.523-528

Wagner, S. and Breiteneder, H. (2002). The latex-fruit syndrome. Biochemical Society Transactions, Vol.6, pp.935-940
Whereas some “microarray” or “bioinformatics” scientists among us may have been criticized as doing “cataloging research”, the majority of us believe that we are sincerely exploring new scientific and technological systems to benefit human health, human food and animal feed production, and environmental protections. Indeed, we are humbled by the complexity, extent and beauty of cross-talks in various biological systems; on the other hand, we are becoming more educated and are able to start addressing honestly and skillfully the various important issues concerning translational medicine, global agriculture, and the environment. The two volumes of this book present a series of high-quality research or review articles in a timely fashion to this emerging research field of our scientific community.

How to reference
In order to correctly reference this scholarly work, feel free to copy and paste the following:

José Carlos Jiménez-López, María Isabel Rodríguez-García and Juan de Dios Alche (2011). Systematic and Phylogenetic Analysis of the Ole e 1 Pollen Protein Family Members in Plants, Systems and Computational Biology - Bioinformatics and Computational Modeling, Prof. Ning-Sun Yang (Ed.), ISBN: 978-953-307-875-5, InTech, Available from: http://www.intechopen.com/books/systems-and-computational-biology-bioinformatics-and-computational-modeling/systematic-and-phylogenetic-analysis-of-the-ole-e-1-pollen-protein-family-members-in-plants