

PROF. ZHONGSHAN GAO (Orcid ID : 0000-0001-6944-9955)

Article type : Original Article: Experimental Allergy and Immunology

Identification of a 62 kDa major allergen from *Artemisia* pollen as a putative galactose oxidase

Wanyi Fu^{1,2*}, Zhongshan Gao^{1,2*§}, Ling Gao^{2*}, Jing Jin², Meiling Liu³, Yuemei Sun^{4§},
Shandong Wu¹, Lingying Wu³, Hongshan Ma³, Yimin Dong³, Xuefeng Wang³, Biyuan Gao⁵,
Huiying Wang⁶, Jaap H. Akkerdaas⁷, Serge A. Versteeg⁷, Ronald van Ree⁷

¹Allergy Research Center, Zhejiang University, Hangzhou, 310058, China.

²Department of Horticulture, College of Agriculture and Biotechnology, Zhejiang University, Hangzhou, 310058, China.

³Department of Allergy, The Third People's Hospital of Datong, Shanxi, 037008, China

⁴Department of Allergy, Yu Huang Ding Hospital, Yantai, Shandong, China.

⁵ Research Department, Hangzhou Aileji Biotech Ltd, Hangzhou, China.

⁶Department of Allergy, the Second Affiliated Hospital, School of Medicine, Zhejiang University, Hangzhou, 310013, China.

⁷Academic Medical Center, University of Amsterdam,
Department of Experimental Immunology and Department of Otorhinolaryngology,
Meibergdreef 9, 1105 AZ Amsterdam, the Netherlands.

* These authors contributed equally to this research.

This article has been accepted for publication and undergone full peer review but has not been through the copyediting, typesetting, pagination and proofreading process, which may lead to differences between this version and the Version of Record. Please cite this article as doi: 10.1111/all.13375

This article is protected by copyright. All rights reserved.

§ Corresponding authors: Dr. Zhongshan Gao and Dr. Yuemei Sun, Department of Horticulture, College of Agriculture and Biotechnology, Zhejiang University, Hangzhou, 310058, China. Email: gaozhongshan@zju.edu.cn; sunym1963@sina.com

Abstract

Background: Around 20 years ago, a 60-70 kDa protein was reported as a major allergen of mugwort (*Artemisia vulgaris*) pollen. This study was to identify and characterize its molecular properties.

Methods: Sera from 113 Chinese and 20 Dutch *Artemisia* allergic/sensitized subjects (and pools thereof) were used to identify the 60-70 kDa allergen. Pollen extracts of seven *Artemisia* species were compared by immunoblotting. Transcriptomics and proteomics (mass spectrometry) of *A. annua* pollen were used to identify the putative 60-70 kDa *Artemisia* allergen. Both the natural purified and recombinant allergens were evaluated for IgE reactivity by ImmunoCAP. Fourteen Chinese *Artemisia* allergic patients were tested intradermally with purified natural allergen.

Results: Immunoblots revealed two major bands at 12 and 25 kDa, and a weak band at 70 kDa for all seven *Artemisia* species. Using a combined transcriptomic and proteomic approach, the high molecular mass allergen in *A. annua* pollen was shown to be a 62 kDa putative galactose oxidase, with a putative N-glycosylation site. More than 94% of *Artemisia* pollen allergic patients had IgE response to this allergen. Although recognition of a non-glycosylated recombinant version was only confirmed in a minority (16%) and at much lower IgE levels, this discrepancy cannot be explained simply by reactivity to the carbohydrate moiety on the natural allergen. Intradermal testing with the natural allergen was positive in five out of nine sensitized patients.

Conclusions: The previously reported 60-70 kDa allergen of *Artemisia* pollen is most likely a 62 kD putative galactose oxidase here designated Art an 7.

Keywords: allergen; *Artemisia annua*; galactose oxidase; proteomics; transcriptomics

Introduction

Artemisia pollens have been recognized as one of the major causes for late summer and autumn allergic respiratory disease in the northern hemisphere (1,2). *Artemisia vulgaris*, mugwort, is one of the best studied weed species in Europe, with six of its seven reported allergens been listed in the IUIS allergen nomenclature database (www.allergen.org). However, the first identified “Art v 60 kDa” still has, surprisingly, no official name due to the lack of molecular information (3,4).

Artemisia pollen sensitization among Chinese pollinosis patients ranges from 58% in the North (high pollen exposure) to only 3% in the South (low pollen exposure) (2,5,6). The dominant *A. vulgaris* species in Northwestern and Central Europe also grows in Western China, but in the densely populated of Northern and Eastern areas, one of the major allergic *Artemisia* species is sweet wormwood (*A. annua*), the source of the antimalarial drug artemisinin (7). *A. annua* was the first allergenic weed pollen reported in Beijing (8), and four other *Artemisia* species (*argyi*, *capillaris*, *lavandulifolia*, *sieversiana*) have been recognized as major sources of allergenic pollen (9).

Accepted Article

In China, five allergens have been identified in *A. annua*, *A. sieversiana* and *A. argyi* pollen extracts using immunoblotting, their estimated molecular mass being 10-12kDa, 16kDa, 20-26kDa, 43kDa and 62 kDa (10-12). Until now, only two *Artemisia* pollen allergens (Art v 1 and Art v 3) have been confirmed using commercially available components (13). The original 62 kDa allergen was not determined for its amino acids sequence and protein type (12). In Europe, about 20 years ago, a 60-70 kDa allergen (Art v 60 kDa) was reported as a monomeric acidic glycoprotein (4,14) with a high sensitization rate of 73% - 92% in mugwort allergic patients (4,14-16). A blocked N-terminal and difficulties in trypsin digestion hampered molecular identification of this allergen at that time, and only a few amino acids of a nonapeptide sequence were identified (4,14). Art v 60 kDa cross-reactive allergens in the oral allergy syndrome to several pollens, vegetables and fruits have also been reported (4,17). Apart from homologous proteins, broad cross-reactivity is due to carbohydrate determinants (CCDs) (18,19).

Despite its potential role as a major weed pollen allergen, the complete amino acid sequence, functional properties, and type of glycosylation of the 60-70 kDa *Artemisia* protein has so far not been characterized. Recent RNA-sequencing techniques and mass spectrometry facilitate gene and protein characterization in specific tissues (20). Combined transcriptomic and immune-proteomic approaches have been applied to storage mite, short ragweed and feverfew pollens to identify novel allergens (21-23). In this research, using a similar approach, we identified the *Artemisia* pollen allergen as a 62 kDa putative galactose oxidase protein, now designated Art an 7 by IUIS allergen nomenclature subcommittee.

Materials and Methods

Patients

In a total of 113 patients (58 male, 55 female), 10 have been reported previously (13). The patients were all recruited between August 2011 and January 2017 in Zhejiang, Shanxi and Shandong provinces in China, based on a convincing history of *Artemisia* pollen allergy (allergic rhinitis, conjunctivitis, asthma, and/or atopic dermatitis), supported by matching sensitization shown by mugwort pollen (w6) ImmunoCAP > 0.35 kU_A/L (Thermo Fisher Scientific, Uppsala, Sweden). Seven non-atopic individuals (N1-N5, N11 and N12) and five mite-allergic patients (N6-N10, sIgE (d1 or d2) > 100 kU_A/L) were recruited as control groups. Specific IgE to two mugwort allergen components, Art v 1 and Art v 3 (w231, w233), was determined for most enrolled patients, but positivity was not an inclusion criterion (**Table 1**). Written informed consent was obtained from all participants and the study was approved by the ethics committee of the Second Affiliated Hospital, College of Medicine, Zhejiang University (authorization No. 2011-R-1, 2013-67-087), in collaboration with The Third People's Hospital of Datong, Shanxi (authorization No. 2015-001) and Yu Huang Ding Hospital, Yantai, Shandong (authorization No. 2015-164).

Sera with specific IgE against mugwort pollen (n=20) from a control serum bank, collected between 1999 and 2004 at the Sanquin Blood Supply Foundation (Amsterdam, The Netherlands), were included in the serological analyses (**Table 2**). No clinical information was available for these subjects due to the fully anonymous nature of the control serum bank.

Immunoblotting

For IgE immunoblot analysis, pollen extracts of seven *Artemisia* species were separated by 12% acrylamide gels and electroblotted (100 V, 350 mA, 1 h) onto PVDF membranes (0.2 μ m, Bio-Rad, Hercules, California, USA) by mini Trans-Blot electrophoretic transfer cell (Bio-Rad). PVDF membranes were blocked with 5% (w/v) skimmed milk, and incubated with a serum pool CN of 35 mugwort allergic patients (Table 1). Bound IgE was detected using goat anti-human IgE coupled with HRP (1:3000 dilution, Thermo Fisher Scientific), then visualized using Clarity western ECL blot substrate (Bio-Rad), and imaged with the ChemiDoc system (Bio-Rad). *A. annua* pollen extract blots were further probed with 15 individual patient sera, and a negative control pool of sera from five non-atopic individuals (N1-N5).

Pollen extraction, transcriptome and proteome

Pollen of six dominant *Artemisia* species in China were collected from Shanxi (*annua*, *argyi*, *capillaris*, *gmelinii*, *sieversiana*) and Zhejiang provinces (*lavandulifolia*), and *A. vulgaris* pollen purchased from Allergopharma (Reinbek, Germany). Protein extracts of pollen were prepared by suspending 2 g of pollen grains in 35 ml PBS buffer (0.14 M NaCl, 2.7 mM KCl, 7.8 mM Na₂HPO₄, 1.5 mM KH₂PO₄) and shaking for 12 h at 4°C. Extracts were centrifuged at 10,000g for 10 min at 4°C, filtered through 0.22 μ m filters (Millipore, Merck KGaA, Darmstadt, Germany), and the protein concentration determined by BCA. Crude pollen extracts were analyzed by SDS-PAGE and visualized by staining with Coomassie brilliant blue R-250.

RNA and protein samples of *A. annua* were prepared as described previously (20). RNA-seq was obtained on the platform of Illumina HiSeq™ 2000 and peptide mass by Q-EXACTIVE (Thermo Fisher, USA). Clean sequence data (5.8 Gb) and 317,291 spectra were generated for analysis (BGI, Shenzhen, China).

Transcriptome sequencing data were obtained by using the SOAP de novo assembly program and the short-read Trinity assembly program, with annotations generated by alignment to NR, Swiss-Prot, COG, KEGG and GO databases (**Figures S1, Figure S2, Figure S3**). Spectra were searched against our *A. annua* transcriptome using the Mascot software, with peptide and fragment mass tolerance set to 20 ppm and 0.6 Da, respectively.

Identification, purification and expression of a 60-70 kDa allergen

The band at about 70 kDa from *A. annua* (AN70), recognized on immunoblot by an *Artemisia*-allergic patient serum pool, was analyzed by LC-MS (Thermo Fisher, LTQ-Velos) and the peptide sequences obtained were aligned with the *A. annua* proteome. Complete open reading frames (ORFs) for the AN70 allergen were amplified from the cDNA of seven *Artemisia* species. A construct was created by cloning the AN70 gene (without the *A. annua* signal sequence) into plasmid pSUMO3 vector as described by the manufacturer (LifeSensors) and expressed in *E. coli* Rosetta 2 (DE3) pLysS. The cell pellet was frozen overnight at -20°C, thawed and re-suspended to one-tenth of the original culture volume. After sonication and pelleting of the insoluble matter by centrifugation, the protein was solubilized for purification in three steps. Firstly, the fractions containing rAN70 were pooled, concentrated and digested with SUMO Protease 2 (LifeSensors) as described by the manufacturer. After digestion, rAN70 was purified by affinity chromatography (AC) using a HisTrap column. The final purification step was Gelfiltration (GF), using a HiLoad 16/60 Superdex 75 pg column. The rAN70-containing fractions were pooled, concentrated, and the purity assessed by SDS-PAGE.

Natural AN70 allergen was purified from *A. annua* pollen extract using anion exchange (HiPrep 16/10 DEAE EF column) and molecular sieve (Superdex 200 Increase 10/300GL) chromatography (GE Healthcare, USA). The purity was assessed by SDS-PAGE, mass spectrometry and immunoblots. The identity was confirmed by LC-MS, and the molecular mass was established by MALDI-TOF-MS (Applied Biosystem, USA). Glycoprotein of *A. annua* pollen extract and of the purified AN70 allergen was detected using a PAS staining kit (Goodbio Technology, Wuhan, China). For deglycosylation, immunoblots were treated with periodate (24). Primary deduced amino acid sequence was used to search the homologous template with BLAST and HHblits against the SWISS-MODEL template library to infer its function and build a 3D model (25). The GlycoMod tool was used to predict the possible oligosaccharide structures that occur on protein from trypsin digested peptide masses (26).

Galactose oxidase activity

Enzymatic oxidation of galactose was measured using a colorimetric HRP-ABTS-coupled assay. To 90µl buffer containing 0.3 M D-galactose, 30 U/ml HRP, 0.1 M NaH₂PO₄ and 0.25 mg/ml ABTS, 10 µl Art an 7 (100 µg/ml) was added. Galactose oxidase (100 µg/ml, from *Dactylium dendroides* (Sigma, USA) was used as positive control at the same concentration. The absorbance at 420 nm was recorded at 30°C for 4 min.

Serological evaluation

IgE reactivity to natural and recombinant allergen was assessed by ImmunoCAP using biotinylated allergen and streptavidin-conjugated ImmunoCAPs (ThermoFisher Scientific) following the manufacturer's instructions. To further characterize IgE recognition of the natural purified allergen, ImmunoCAP inhibition was also performed using the natural (100

Accepted Article

µg/ml) and recombinant (67.5 µg/ml) Art an 7 allergen, *Artemisia* extracts (2.5 mg/ml) and the CCD-marker bromelain (1 mg/ml) as inhibitors, with the serum incubated with inhibitor at room temperature for 1 h, prior to evaluation by ImmunoCAP loaded with the natural purified Art an 7 allergen.

Intradermal skin testing

Fourteen patients (P54, P101-113 in Table 1) underwent intradermal skin testing using 15 µl of *A. annua* pollen extract (0.5µg/ml) and of purified natural Art an 7 allergen (0.5µg/ml), with histamine (10µg/ml) as the positive and saline solution the negative controls. The skin test was considered positive if the wheal diameter was more than 5 mm. Two non-atopic subjects (N11, N12) were tested as a negative control, and skin tests were performed outside the mugwort pollen season.

Results

Chinese patients' characteristics

Amongst the 113 *Artemisia* pollen allergic patients from China (**Table 1**), the most commonly reported symptom was rhinitis (95%), followed by conjunctivitis (58%), asthma (56%), and urticaria (22%). Seasonal allergy symptoms generally occurred from the beginning of August to early October, in line with the local *Artemisia* pollen season. The median specific IgE against *A. vulgaris* pollen extract, Art v 1 and Art v 3 were 40.0, 9.8 and 15.2 kUA/L, respectively, with 78% of the patients having positive responses to Art v 1, 63% to Art v 3 and 54% to both.

Immunoblot analyses of different *Artemisia* species

Pollen extracts of seven *Artemisia* species separated by SDS-PAGE (**Figure 1A**) were transferred to a PVDF membrane for immunoblotting with a serum pool of 35 sera (Table 1: P1-P35) with sensitization to both Art v 1 and Art v 3. Two major bands consistent with Art v 1 and Art v 3 sensitization were observed for all species (**Figure 1B**). A weak but significant band was also detected at around 70 kDa (AN70, previously referred to Art v 60 kDa) for all seven species (**Figure 1B**). *A. annua* pollen extract immunoblots were then probed with 15 individual sera (see **Table 1**), of which one (P90) was not sensitized to either Art v 1 or Art v 3. Recognition of the 70 kDa band (**Figure 1C**) was weak to moderate in all 15 sera.

The 60-70k Da allergens as putative galactose oxidase

The ~70 kDa IgE-binding band of the *A. annua* pollen extract was analyzed by LC-MS after in-gel trypsin digestion. Peptide spectrum matching gave the highest sequence coverage (69.5%) to Unigene 18656 (**Table S1**, with very high protein score), an acidic protein, annotated as a putative galactose oxidase in *Helianthus annuus* (NCBI: OTG34910.1). As Art v 60 kDa has previously been reported to be an acidic protein, we selected the top-matched acidic protein (Unigene 18656) as a candidate. cDNA cloning of the full sequences of the homologous protein from *A. annua* and the other six *Artemisia* species, gave translated protein sequences with variation at 36 positions (**Figure S4**, NCBI Accession No.: KY428925-KY428931), Art an 7 was 98% identical to Art v 7. The protein precursor had 594 amino acids, the first 23 being a predicted signal peptide. N terminal sequencing failed because the N-terminus was blocked. Searching the deduced amino acid sequence of Art an 7 against the SMTL (SWISS-MODEL template library) with BLAST, the allergen was identified as galactose oxidase homologous protein, with the typical kelch, beta-propeller motif and Cu ion ligand binding sites (**Figure S5**). However, nArt an 7 had no enzymatic activity on

D-galactose (**Figure S6**), which may be inactivated on purification. The calculated molecular mass of the mature Art an 7 protein was 62027 Da with a *pI* of 4.75. Based on this analysis, we purified this protein from *A. annua* pollen extract (**Figure 2A**) using anion exchange chromatography and molecular sieve, and subsequently verified its sequence as a putative galactose oxidase and an NLT motif as a potential N-glycosylation site (**Figure 2B**). The main peak mass of the purified protein was 61954.35 Da, 67Da less to the calculated molecular mass (**Figure 2C**), followed by small peaks between 62 and 68 kDa. The assessed purity was over 90% by SDS-PAGE, mass spectrometry and confirmed by immunoblot (Figure 4A, lane 3). Using the GlycoMod tool, one tryptic peptide was found to have a mass of 3313.6 Da, matching a predicted peptide KGTSGWEDATEANLTPVLYTPYKPMGQR (theoretical mass 3109.5 Da) with a HexNac1 (theoretical mass of 203.0Da) linked to the putative glycosylation site.

Immunological characterization of Art an 7

IgE against the natural purified putative galactose oxidase, designated wormwood Art an 7 by IUIS, was measured by ImmunoCAP in sera of 113 Chinese *Artemisia* allergic patients (**Table 1**) resulting in remarkably high recognition (94%). Negative controls of non-allergic sera (N1-N5) and mite-allergic sera (N6-N10) were negative to both natural and recombinant Art an 7. For the 113 Chinese patients, the median nArt an 7-specific IgE titer was 5.4 kU_A/L, lower than nArt v 1 (9.8 kU_A/L) and nArt v 3 (15.2 kU_A/L) (**Figure 3A**). IgE responses to nArt an 7 was moderately correlated to the IgE titers against mugwort extract (**Figure 3B**), but not to those against Art v 1 or Art v 3 (**Figure 3C and D**). Similar sensitization (95%) was found in 20 Dutch mugwort-sensitized subjects, with a lower median sIgE titer (1.98 kU_A/L) (**Table 2**). Surprisingly, only 4% of the samples from *Artemisia*-sensitized Chinese (2/54) but 50% (10/20) from the Dutch subjects bound to the rArt an 7, in most cases with significantly

lower IgE titers. Both purified Art an 7 and extract showed the presence of glycoprotein by PAS staining, and deglycosylation did not affect much of its IgE binding to Art an 7 (**Figure 4A**). To investigate whether this striking difference could be explained by binding to the glycan moiety on natural Art an 7, a selection (21/113) of the Chinese and all Dutch sera were tested for IgE against the CCD marker protein bromelain (nAna c 2). Only 4/21 Chinese sera were weakly sensitized to CCD ($< 1 \text{ kU}_A/\text{L}$), but the IgE titers against nArt an 7 were 9- to 27-fold higher. Amongst the Dutch sera, 8/20 sera were positive for CCD, with high IgE titers in some cases, indicating that IgE binding to CCD may partly explain the IgE reactivity to nArt an 7 for a sub-group of the Dutch sera. A pool of four CCD-positive Dutch sera (see **Table 2**) was then subjected to nArt an 7 ImmunoCAP inhibition with bromelain (**Figure 4B**). Binding to nArt an 7 was equally inhibited by mugwort extract, nArt an 7 and bromelain (around 80%), while there was no significant inhibition by recombinant Art an 7 (around 10%). In a pool of Dutch sera without IgE against CCD (see **Table 2**) was subjected to the same protocol, bromelain did not show any inhibition, but there was still only borderline inhibition of rArt an 7, around 10%. For the pool of Chinese sera used for immunoblotting (P1-P35; Table 1), there was around 90% inhibition with nArt an 7, and similarly only borderline inhibition with bromelain (around 10%). IgE recognition of nArt an 7 in an immunoblot with nArt an 7 and with *A. annua* pollen extract treated with periodate was not significantly affected, whereas recognition of two other allergens disappeared completely. Together, these data illustrate that the lack of IgE recognition of rArt an 7 cannot be accounted for by the absence of CCD, although in some (Dutch) sera CCD does appear to play a role in this recognition.

Accepted Article

Finally, the allergenicity of Art an 7 was assessed by intradermal skin testing in 14 Chinese patients, seven gave a positive reaction (**Table 3**). All except one (0.23 kU_A/L) had a positive ImmunoCAP to nArt an 7 (> 0.35 kU_A/L), and of the seven patients with negative intradermal skin tests, two had low sensitization to nArt an 7 (0.39 and 0.59 kU_A/L, respectively).

Discussion

Using state-of-the-art transcriptomics and proteomics, we have been able to identify the major *Artemisia annua* pollen allergen similar to “Art v 60 kDa”, reported over two decades ago, as a 62 kDa putative galactose oxidase in this study. Protein identity has been assured by experimental whole protein and digested peptide mass spectrum. The blocked N-terminal of nArt an 7 prevented direct sequencing, confirming the previous report on Art v 60 kDa (4). Rabbit polyclonal antibodies to three synthesized peptides, including one of the first 14 N terminal amino acids (APDITDGGDKPGPE), were all reactive to the targeted nArt an 7 in western blots (**Figure S7**). From our deduced mature Art v 7 amino acid sequence, the numbers of the three amino acids (Ala, Ser and His) matched that previously reported, i.e. 47, 27 and 8, respectively (4), but we did not find the reported complete ‘APGPSXXNP’ nonapeptide in nArt an 7, even in our protein database. Molecular mass estimation of Art v 60 kDa in the past has ranged from 60 to 70 kDa, our mass determination revealed that the pure protein mass was 62 kDa, minor glycosylation resulting in extension of a few hundred dalton to 6 kDa. The theoretical isoelectric point of this protein was 4.75, close to the previously reported value of 4.4 (14). Our study confirms earlier claims that the group 7 *Artemisia* pollen allergen is a major allergen, acidic and glycoprotein (4,14). Though not abundant as Art an 1 and Art an 3 in pollen, its prevalence in mugwort allergic patients is the highest. Moreover, we have demonstrated in vivo biological activity by intradermal skin testing. A weakness of the present study is the poor performance of the recombinant

Accepted Article

version of the *A. annua* group 7 allergen, rArt an 7. With galactose oxidase being a glycoprotein, we investigated whether recognition of nArt an 7 could be explained by CCD only. This was ruled out by three lines of evidence: 1) most serum samples with IgE against nArt an 7 from both China and The Netherlands had no IgE against the CCD marker bromelain; 2) binding of a Chinese serum pool to nArt an 7 could not be inhibited with bromelain; and 3) binding of the same serum pool to nArt an 7 on immunoblot was not affected by periodate treatment. However, IgE binding to CCD on nArt an 7 was demonstrated for a subgroup of CCD-positive Dutch sera, but, overall, CCD does not fully explain the discrepancy between the IgE responses to the natural and the recombinant allergen. By SDS-PAGE and mass spectra, the apparent purity of nArt an 7 was high, and immunoblot analyses and inhibitions did not point to IgE binding to other allergens. Therefore, we consider it unlikely that impurities explain this difference in the IgE responses. Moreover, it cannot be ignored that the rArt an 7 does bind IgE, albeit with low potency, to some mugwort allergic sera, but not to mite allergic sera of very high specific IgE titers (**Table 1**, N6-N10).

Art an 7 is the highest molecular weight allergen found in pollen to date. Galactose oxidase is a monomeric protein containing a single copper ion, which catalyzes the oxidation of D-galactose (27). Galactose is the main sugar in pollen, and it was discovered forming a backbone linked to a proline-rich domain in natural Art v 1 allergen (28,29). We found a NLT motif and mass spectrometry evidence of HexNac1 glycosylation, which is conserved in seven *Artemisia* species (**Figure S4**) and predicted on the surface (**Figure S5**). Our finding of one HexNac is uncommon in plant glycoprotein, as the conserved structure of N-linked glycan usually has a core region consisting of two N-acetylglucosamine residues (30). More research is needed to identify the composition of the glycan and impact on allergenicity of

Art an 7. To date, IgE with high cross-reactivity of pollen galactose oxidase has only been found in the pollen of short ragweed, a plant closely related to *Artemisia* species, by Proteomics-Informed-by-Transcriptomics (21). Galactose oxidase is likely to be another cross-reactive allergen between ragweed and mugwort. Apart from the high prevalence of sensitization, earlier reports on “Art v 60 kDa” showed that the allergen has a broad spectrum of cross-reactivity to several other pollens and to multiple foods, suggesting that it may be involved in the mugwort pollen-food syndrome (4,14). Searching for homologous sequences in vegetables, fruits and spices (**Table S2**), the highest identity (74%) was to *Helianthus annuus*, another allergen resource in the same family as *Artemisia* spp., with identity to other species from 42% to 52%. This study may promote further research into the role of putative galactose oxidases in food allergic reactions, sometimes severe, associated with mugwort pollen allergy.

We identified 2,035 proteins were identified from the *A. annua* pollen transcriptome and proteome, and seven were confirmed with high homology to reported mugwort allergens (**Figure S8**). Putative Art an 1 was highly similar (94% -98% identity) to Art v 1 isoforms (31), in contrast, the amino sequence in Art an 3 differed from Art v 3 at 15 positions (84% identity). This difference in identity may affect the specificity of the immune-reactivity of *A. annua* pollen. The other five homologous putative allergens had identities greater than 95% to European mugwort (*A. vulgaris*) allergens (Art v 2, 4, 5 and 6). Gene cloning and transcript analysis of major allergens among the main Chinese local *Artemisia* species are underway for better insight into sequence variation and for their immunological assessment.

In conclusion, we identified a previously reported allergen from *Artemisia* pollen as a putative galactose oxidase and confirmed its role as a major allergen in *Artemisia* pollen allergic patients. The natural molecule can be used for component-resolved diagnosis, and also to investigate whether galactose oxidase is a missing link in mugwort pollen-associated food allergies.

Author contribution

ZSG, RVR, YMS and HSM conceived and designed the study. WYF, LG, JJ and BYG collected pollens, performed the gene cloning, natural protein purification and transcriptome /proteome analysis. MLL, YMS, LYW, YMD, XFW and HYW were involved in the clinical study of patients. WYF, SDW and BYG conducted ImmunoCAP test and data analysis. SAV and JHA did the expression of the recombinant proteins and ImmunoCAP tests and inhibition experiments. WYF, ZSG and RvR drafted the manuscript in close collaboration with all co-authors.

Acknowledgements

This study was supported by a grant from Hangzhou Aileji BioTech, China. Ms Xiaohan Zhu assisted in purification of natural allergen. Ms Hongli Chen, Jing Yang, Wenjun Wang, Xiaojin Du from Datong 3rd People's Hospital are acknowledged for their contribution in the sera collection. Ms Yi Liu and Mr. Zhaowei Yang, Mr. Xiang Zhou helped to test some CAPs of mugwort. We thank Dr. Richard E. Goodman, Chairman of the Executive Committee of WHO/IUIS Allergen Nomenclature Sub-committee for comments about this new allergen.

Disclosure of potential conflict of interest

ZS Gao and WY Fu have received a grant from Hangzhou Aileji Biotech Ltd (2016-R1), China. R. van Ree has received payment for lectures from Phadia. A patent for Artemisia pollen galactose oxidase has been applied for in China (201710097440.7 pending) by ZS Gao, WY Fu, BY Gao and L Gao. The remaining authors declare that they have no relevant conflicts of interest.

References

1. Movérare R, Larsson H, Carlsson R, Holmquist I. Mugwort-sensitized individuals from North Europe, South Europe and North America show different IgE reactivity patterns. *Int Arch Allergy Immunol*. 2011;154(2):164–72.
2. Tang R, Sun JL, Yin J, Li Z. Artemisia Allergy Research in China. *Biomed Res Int*. 2015;2015:179426. DOI:10.1155/2015/179426
3. Gadermaier G, Hauser M, Ferreira F. Allergens of weed pollen: an overview on recombinant and natural molecules. *Methods*. 2014;66(1):55–66.
4. Heiss S, Fischer S, Muller WD, Weber B, Hirschwehr R, Spitzauer S, et al. Identification of a 60 kd cross-reactive allergen in pollen and plant-derived food. *J Allergy Clin Immunol*. 1996;98:938–47.
5. Zheng YW, Li J, Lai XX, Zhao DY, Liu XF, Lin XP, et al. Allergen micro-array detection of specific IgE-reactivity in Chinese allergy patients. *Chin Med J (Engl)*. 2011;124(24):4350–4.
6. Hao GD, Zheng YW, Gjesing B, Kong XA, Wang JY, Song ZJ, et al. Prevalence of sensitization to weed pollens of *Humulus scandens*, *Artemisia vulgaris*, and *Ambrosia artemisiifolia* in northern China. *J Zhejiang Univ Sci B*. 2013;14(3):240–6.

7. Graham IA, Besser K, Blumer S, Branigan CA, Czechowski T, Elias L, et al. The genetic map of *Artemisia annua* L. identifies loci affecting yield of the antimalarial drug artemisinin. *Science*. 2010, 327: 328–331.
8. Leng X, Ye ST. An investigation on in vivo allergenicity of *Artemisia annua* leaves and stems. *Asian Pac J Allergy Immunol*. 1987;5(2):125–8.
9. Qiao BS, Wang LL, Yin J, Niu TP, Bao JN, Liu GH, et al. Color Atlas of air-borne pollens and plants in China Peking Union Med. Coll Press Beijing. 2004; (in Chinese).
10. Yang H, Liu ZG, Han QG, Hou SB, Liang GZ. Analysis of allergen composition in *Artemisia argyi* and *Artemisia apiacea* pollen. *Imunol J*. 2004;20(2):120–3.
11. Yao LN. pollen survey by volumetric sampler in Beijing city and cross-reactivity research between ragweed and mugwort in vitro. PhD thesis, Peking Union Medical College, 2009; (in Chinese, English abstract).
12. Li XY, Xiao X, Sun HZ, He SH, Yang PC, Liu ZG. Identification and purification of major allergens in *Artemisia sieversiana* pollen. *Chinese J Immunol*. 2014;7:913–6 (in Chinese, English abstract).
13. Gao ZS, Yang ZW, Wu SD, Wang HY, Liu ML, Mao WL, et al. Peach allergy in China: a dominant role for mugwort pollen lipid transfer protein as a primary sensitizer. *J Allergy Clin Immunol*. 2013;131(1):223–4.
14. de la Hoz F, Polo F, Del Prado JM, Selles JG, Lombardero M and Carreira J. Purification of Art v I, a relevant allergen of *Artemisia vulgaris* pollen. *Mol Immunol*. 1990;27(7):651–7.
15. Lombardero M, García-Sellés FJ, Polo F, Jimeno L, Chamorro MJ, García-Casado G, et al. Prevalence of sensitization to *Artemisia* allergens Art v 1, Art v 3 and Art v 60 kDa. Cross-reactivity among Art v 3 and other relevant lipid-transfer protein allergens. *Clin Exp Allergy*. 2004;34(9):1415–21.

- Accepted Article
16. Oberhuber C, Ma Y, Wopfner N, Gadermaier G, Dedic A, Niggemann B, et al. Prevalence of IgE-binding to Art v 1, Art v 4 and Amb a 1 in mugwort-allergic patients. *Int Arch Allergy Immunol*, 2008; 145:94-101.
 17. Bauer L, Ebner C, Hirschwehr R, Wüthrich B, Pichler C, Fritsch R, et al. IgE cross-reactivity between birch pollen, mugwort pollen and celery is due to at least three distinct cross-reacting allergens: Immunoblot investigation of the birch-mugwort-celery syndrome. *Clin Exp Allergy*. 1996;26(10):1161–70.
 18. Aalberse RC, Koshte V, Clemens JG. Immunoglobulin E antibodies that crossreact with vegetable foods, pollen, and hymenoptera venom. *J Allergy Clin Immunol*. 1981;68(5):356-64.
 19. van Ree R. Carbohydrate epitopes and their relevance for the diagnosis and treatment of allergic diseases. *Int Arch Allergy Immunol*. 2002;129(3):189-97.
 20. Wu HX, Jia HM, Ma XW, Wang SB, Yao QS, Xu WT, et al. Transcriptome and proteomic analysis of mango (*Mangifera indica* Linn) fruits. *J Proteomics*. 2014;105:19–30.
 21. Floch VB, Mignon ML, Bouley J, Groeme R, Jain K, Baron-Bodo V, et al. Identification of Novel Short Ragweed Pollen Allergens Using Combined Transcriptomic and Immunoproteomic Approaches. *PLoS One*. 2015;10(8):e0136258.
 22. Cui Y, Yu L, Teng F, Zhang C, Wang, N, Yang, L, Zhou Y. Transcriptomic/proteomic identification of allergens in the mite *tyrophagus putrescentiae*. *Allergy*. 2016; 71(11): 1635-1639.
 23. Pablos I, Eichhorn S, Briza P, Asam C, Gartner U, Wilf M et al. Proteomic profiling of the weed feverfew, a neglected pollen allergen source. *Scientific Reports*, 2017; 7(1):6049.
 24. Borghesan F, Mistrello G, Amato Et Al S, Amato S, Villalta D, Asero R.
Mugwort-fennel-allergy-syndrome associated with sensitization to an allergen homologous to Api g 5. *Eur Ann Allergy Clin Immunol*. 2013;45(4):130–7.
 25. Remmert M, Biegert A, Hauser A and Soding J. HHblits: lightning-fast iterative protein sequence searching by HMMHMM alignment. *Nat Methods*, (2012)9, 173-175.

26. Cooper CA, Gasteiger E, Packer N. GlycoMod - A software Tool for Determining Glycosylation Compositions from Mass Spectrometric Data Proteomics 2001, 1:340-349.
27. Parikka K, Master E, Tenkanen M. Oxidation with galactose oxidase: Multifunctional enzymatic catalysis. J Mol Catal B Enzym. 2015;120:47–59.
28. Leonard R, Petersen BO, Himly M, Kaar W, Wopfner N, Kolarich D, et al. Two novel types of O-glycans on the mugwort pollen allergen Art v 1 and their role in antibody binding. J Biol Chem. 2005;280(9):7932–40.
29. Xie N and Taylor CM, Synthesis of Oligomers of β -l-Arabinofuranosides of (4R)-4-Hydroxy-l-proline Relevant to the Mugwort Pollen Allergen, Art v 1 J. Org. Chem., 2014; 79 (16):7459–67.
30. Homann A, Schramm G, Jappe U. Glycans and glycan-specific IgE in clinical and molecular allergology: Sensitization, diagnostics, and clinical symptoms. J Allergy Clin Immunol, 2017. 140(2): 356-36831. Dedic A, Gadermaier G, Vogel L, Ebner C, Vieths S, Ferreira F, Egger M. Immune recognition of novel isoforms and domains of the mugwort pollen major allergen art v 1. Molecular Immunology, 2009; 46(3): 416-421.

Figure Legends

Figure 1. Allergen identification from seven Artemisia pollen extracts by immunoblotting.

A: Protein extracts were separated by SDS-PAGE

B: Immunoblots by serum pool CN using seven Artemisia pollen extracts lane 1-7 (*A. annua*, *-argyi*, *-capillaris*, *-gmelinii*, *-lavandulifolia*, *-sieversiana*, *-vulgaris*).

C: Immunoblots of individual serum with *A. annua* extract. *A. annua* extract was separated by SDS-PAGE (S), and incubated with sera from 15 mugwort-allergic patients (lane 2-16:

P49-P60, P88-P90). A sera pool from five healthy individuals was used as control (lane N).

Figure 2. nArt an 7 (nAN70) purification, mass spectrometry and peptide matches. A:

SDS-PAGE of the purified Art an 7 (AN70); B: Mature protein sequence, verified peptides by mass spectra fingerprint shown in red and highlighted (NCBI accession KY428925), NLT was a putative N-glycosylation motif; C: mass spectra of the purified Art an 7.

Figure 3. Specific IgE levels to mugwort extract, nArt v 1, nArt v 3 and nArt an 7 determined by ImmunoCAP (A), and correlations of IgE levels between nArt an 7 to allergens (B, C, D).

The positive rate of allergens are displayed below, the median as a solid line, and cutoff (0.35 kU_A/L) as red dashed line. Low ImmunoCAP values (IgE<0.15 kU_A/L) are given as 0.15 kU_A/L to allow a log-scale presentation of the data. The Spearman rank correlation (r) and P value are shown in the lower right corner.

Figure 4. Reactivity of Art an 7 and CCD. A: Purified nArt an70 (lane 1, 1 µg) and *A. annua* pollen extract (lane 2, 50 µg) were stained using the Periodic Acid-Schiff method. Natural (lane 3) and de-glycosylated nArt an 7 (lane 4) were tested by immunoblotting with the CN serum pool. B: Inhibitions of seven Artemisia pollen extracts (2.5 mg/ml), nArt an 7 (100 µg/ml), rArt an 7 (67.5 µg/ml) and bromelain (1 mg/ml) to nArt an 7 with serum pools CCD+, CCD- and CN.

Supplementary materials: Appendices

Appendix 1: Figure S1. NR classification of *A. annua* pollen transcriptome. The species distribution of the result of NR annotation (A) E-value distribution of the result of NR annotation (B), and the similarity distribution of the result of NR annotation (C).

Appendix 2: Figure S2. GO classification analysis of proteins in proteome. On the left are GO functions divided into three ontologies: molecular function, cellular component and biological process. The number of proteins which have the GO function is shown on the right.

Appendix 3: Figure S3 COG function classification of proteins in proteome.

The horizontal ordinate letters on the right stand for function classes of COG, and the vertical coordinates are numbers of proteins of each class.

Appendix 4: Table S1. Top 10 identified proteins in *A. annua* pollen proteome

Appendix 5: Figure S4. Deduced galactose oxidase precursor protein sequences in seven *Artemisa pollens* (NCBI: KY428925-KY428931). Glycosylation site NLT was boxed.

Appendix 6 Figure S5 3D model of Art an 7 protein with Blast and HHBlits against the SWISS-MODEL template 4unm. a, b: Typical kelch beta-propeller structure 3D views from different angel, pink dot is Cu, N-glycosylation site NLT locates on the surface. c: the active site showing ligand of copper and residues.

Appendix 7 Figure S6 Enzymatic activity on D-galactose of nArt an 7 and GAOX from *Dactylium dendroides*

Appendix 8 Figure S7 Reaction of polyclonal antibodies derived from three synthesized peptides from Art an 7 amino acid sequence to naturally purified Art an 7.

Appendix 9: Table S2. Art an 7 homologous sequences in vegetables, fruits and spices with high identity.

Appendix 10: Figure S8. Deduced mature peptides sequences in *A. annua* proteome of seven Art v allergens (*A. vulgaris*) homologous.

Table 1. Chinese patient information (n=113), allergy symptoms, and ImmunoCAP specific IgE values (kU_A/L)

Sera code	Gender	Age	Symptom	sIgE (kU _A /L)						Usage
				mugwort	nArt v 1	nArt v 3	nArt an 7	rArt an 7	nAna c 2	
P1	M	9	AS,C,R	169.00	22.50	147.00	18.90	16.10	ND	Pool
P2	F	15	AS,R	92.60	55.60	58.00	13.30	0.21	ND	Pool
P3	F	9	AS,C,R	171.00	68.40	106.00	20.00	0.09	0.14	Pool
P4	M	21	AS,C,R	53.00	18.10	67.10	10.50	0.07	ND	Pool
P5	M	10	AS,R,E	91.00	89.10	1.30	6.95	0.06	0.75	Pool
P6	F	50	AS,C,R	78.00	69.00	0.36	13.20	0.05	ND	Pool
P7	F	16	AS,C,R,E	145.00	4.21	16.30	15.10	0.04	ND	Pool
P8	F	13	R,E	84.00	16.60	74.90	25.30	0.04	0.07	Pool
P9	M	4	AS,R,E	99.00	1.33	164.50	33.10	0.03	0.16	Pool
P10	M	13	AS,C,R,E	54.70	27.60	45.70	8.93	0.02	0.11	Pool
P11	F	14	AS,C,R	73.90	13.80	8.07	11.10	0.02	0.41	Pool
P12	F	17	AS,R,E	66.30	8.10	73.70	14.30	0.02	0.13	Pool
P13	F	17	AS,C,R	84.90	43.10	64.90	9.51	0.01	0.07	Pool
P14	F	47	AS,R	40.60	14.30	44.50	8.15	0.01	0.36	Pool
P15	M	12	C,R	55.90	60.50	0.44	9.94	0.01	ND	Pool
P16	F	9	C,R	68.90	1.53	25.20	9.46	0.01	0.43	Pool
P17	M	42	AS,R,E	58.00	9.68	74.40	5.99	0.00	ND	Pool
P18	M	12	R	19.00	7.01	22.90	2.08	0.00	ND	Pool
P19	M	12	R,E	40.40	0.39	60.00	10.60	ND	ND	Pool
P20	M	45	AS,C,R	45.30	26.90	32.20	2.08	ND	ND	Pool
P21	M	13	AS,C,R,E	51.90	38.70	32.80	8.01	ND	0.03	Pool
P22	M	13	AS,R	31.80	9.78	29.10	2.87	ND	ND	Pool
P23	F	14	AS,R	59.90	29.00	54.30	3.84	ND	ND	Pool
P24	F	7	C,R	19.50	12.10	12.60	3.00	ND	ND	Pool
P25	M	33	AS	21.60	4.08	20.20	2.26	ND	ND	Pool
P26	F	9	AS,R	31.30	18.80	16.20	7.12	ND	ND	Pool
P27	F	22	AS,R	43.90	40.40	0.42	21.80	ND	ND	Pool
P28	F	44	AS,R,E	11.60	7.38	9.03	2.04	ND	ND	Pool
P29	F	22	AS,R	16.10	8.35	15.60	2.33	ND	ND	Pool
P30	M	20	C,R	27.20	0.97	35.30	1.31	ND	ND	Pool

P31	M	11	C,R	249.00	87.40	176.00	24.90	ND	ND	Pool
P32	M	12	C,R,E	40.00	5.09	69.90	4.33	ND	ND	Pool
P33	F	17	AS,C,R	14.60	5.16	37.30	2.81	ND	ND	Pool
P34	F	44	AS,C,R	22.60	14.80	33.20	4.61	ND	ND	Pool
P35	M	13	C,R,E	44.80	3.41	23.70	17.90	ND	ND	Pool
P36	M	27	AS,C,R	18.30	1.63	25.20	4.06	ND	ND	
P37	M	12	C,R	91.70	3.33	23.50	7.76	ND	ND	
P38	F	7	R	193.00	167.00	31.50	4.66	0.04	ND	
P39	F	16	C,R	239.00	91.00	13.20	6.24	0.01	ND	
P40	M	8	R,E	190.00	2.20	335.00	1.15	0.00	ND	
P41	M	12	R	127.00	2.10	15.80	5.02	0.01	ND	
P42	M	34	AS	23.30	1.14	20.50	3.89	0.01	ND	
P43	M	52	AS,C,R	73.90	0.86	100.00	0.88	0.01	ND	
P44	M	22	C,R	139.00	43.50	131.00	3.03	0.00	ND	
P45	F	44	C,R	47.90	16.50	51.20	0.73	0.01	ND	
P46	F	50	AS,R	31.40	16.20	16.80	3.96	0.00	ND	
P47*	F	22	AS	97.50	27.90	90.10	0.03	0.01	ND	
P48	F	53	AS,C,R	22.70	10.00	21.90	0.84	0.00	ND	
P49	M	12	R	292.00	71.90	200.00	8.42	0.01	ND	IB
P50*	F	26	R	125.70	70.30	130.00	8.85	0.14	ND	IB
P51*	F	42	AS,R	38.80	3.14	36.90	6.65	0.00	ND	IB
P52*	M	18	AS,R	104.10	69.60	101.11	42.70	0.33	ND	IB
P53	F	11	AS,C,R,E	111.00	21.30	61.80	30.60	0.02	ND	IB
P54	F	57	AS,C,R,E	56.10	21.70	5.02	5.70	ND	ND	IB, IT+
P55	M	54	AS,R	121.00	24.20	99.00	20.60	ND	ND	IB
P56	F	23	R	67.20	30.60	0.20	27.70	0.07	ND	IB
P57	M	10	C,R,E	31.40	22.40	0.06	27.30	0.01	ND	IB
P58	M	11	AS,C,R,E	75.00	68.60	0.25	18.20	0.01	ND	IB
P59	F	7	C,R,E	62.00	41.90	0.11	17.10	ND	ND	IB
P60*	F	42	AS,R	18.40	17.30	0.10	1.80	ND	ND	IB
P61*	F	13	R,E	44.70	83.50	0.17	5.93	0.05	ND	
P62	F	22	C,R	49.50	51.40	0.26	6.25	0.00	ND	
P63	F	14	C,R	20.70	20.40	0.08	9.65	0.00	ND	
P64	M	27	AS,R,E	44.50	17.30	0.13	16.30	0.00	ND	
P65	M	25	AS	15.10	11.50	0.05	1.96	0.00	ND	
P66	F	36	R	11.20	9.78	0.05	1.44	0.00	ND	
P67	F	40	C,R	17.30	7.12	0.02	2.92	0.00	ND	
P68	F	8	C,R	5.70	5.56	0.01	1.81	0.01	ND	
P69	M	35	C,R	4.35	4.71	0.01	1.33	ND	ND	
P70	F	33	AS,R,E	26.10	55.60	0.08	4.11	ND	ND	
P71	M	14	AS,R	43.70	44.80	0.13	4.56	ND	ND	
P72	F	14	C,R	30.10	28.90	0.29	3.72	ND	ND	
P73	F	14	AS,R	51.70	28.70	0.11	12.10	ND	ND	
P74*	F	26	AS,E	21.30	21.40	0.13	9.00	ND	ND	
P75*	F	53	AS,C,R	10.70	16.10	0.08	1.52	ND	ND	
P76	F	20	C,R	12.40	13.80	0.07	4.76	ND	ND	
P77	F	31	AS	9.98	10.50	0.02	2.30	ND	ND	
P78	M	15	C,R	13.10	10.20	0.04	4.05	ND	ND	

P79	M	17	C,R	8.80	8.33	0.05	1.81	ND	ND	
P80	F	42	AS,C,R	6.38	6.23	0.06	5.27	ND	ND	
P81	M	11	C,R	8.09	0.06	10.30	5.44	ND	ND	
P82	M	15	AS,C,R	76.90	0.29	84.00	12.00	ND	ND	
P83	M	11	AS,R	8.14	0.12	2.28	15.00	ND	ND	
P84	F	17	R	527.00	0.00	207.00	0.65	0.04	ND	
P85	M	6	C,R	49.20	0.05	34.20	17.10	0.01	ND	
P86	M	13	C,R	7.70	0.04	11.90	6.20	0.00	ND	
P87	F	42	R	9.41	0.24	15.20	0.00	0.00	ND	
P88	M	11	AS,C,R	93.60	0.21	5.09	60.90	0.02	ND	IB
P89	F	12	R	28.10	0.01	0.00	22.00	ND	ND	IB
P90	F	17	C,R	26.20	0.06	0.00	10.20	ND	ND	IB
P91	F	13	C,R	43.50	0.01	0.00	3.14	ND	ND	
P92	F	31	AS,R,E	5.02	0.00	0.00	2.10	ND	ND	
P93	F	52	AS,R	5.56	0.17	0.14	1.33	ND	ND	
P94	F	33	AS,C,R	51.50	0.10	0.01	33.80	ND	ND	
P95*	F	11	R,E	17.30	0.09	0.08	8.59	ND	ND	
P96	M	15	AS,C,R	32.00	0.05	0.00	23.40	ND	ND	
P97	M	10	AS,C,R	23.20	0.03	0.00	3.08	0.54	ND	
P98	F	12	C,R	9.96	0.01	0.00	5.95	0.00	ND	
P99	F	9	C,R	23.10	0.01	0.00	6.41	0.00	ND	
P100*	F	22	R	11.20	0.09	0.09	12.50	0.03	ND	
P101	F	51	AS,R,E	1.29	1.37	0.00	0.03	ND	0.01	IT-
P102	F	23	AS,C,R	94.80	53.10	51.90	10.50	ND	ND	IT+
P103	F	59	C,R	5.63	5.37	0.00	0.40	ND	0.01	IT-
P104	F	61	AS,C,R	79.50	35.60	126.50	2.99	ND	ND	IT+
P105	M	24	C,R	35.70	33.30	8.34	1.00	ND	ND	IT+
P106	M	29	AS,C,R	14.00	0.18	23.00	0.68	ND	0.03	IT-
P107	F	23	R	0.49	0.00	0.04	0.34	ND	0.01	IT-
P108	M	27	C,R	2.56	0.04	0.00	0.08	ND	0.01	IT-
P109	M	30	C,R	45.50	0.61	70.90	0.95	ND	0.18	IT+
P110	M	52	AS,C,R	14.30	1.10	26.60	0.23	ND	0.00	IT+
P111	M	49	AS,C,R	14.60	0.33	21.80	0.39	ND	0.00	IT-
P112	F	41	C,R	3.49	0.27	0.01	0.18	ND	0.12	IT-
P113	M	44	AS,C,R	16.90	4.22	19.60	0.59	ND	0.01	IT+
N1	M	33	N	0.01	ND	ND	0.02	0.04	ND	
N2	M	28	N	0.01	ND	ND	0.01	0.01	ND	
N3	F	26	N	0.00	ND	ND	0.00	0.04	ND	
N4	M	25	N	0.01	ND	ND	0.03	0.06	ND	
N5	M	47	N	0.00	ND	ND	0.00	0.02	ND	
N6	M	4	N	0.05	ND	ND	0.01	0.02	ND	
N7	F	8	N	0.03	ND	ND	0.00	0.01	ND	
N8	F	6	N	0.03	ND	ND	0.00	0.04	ND	
N9	M	7	N	0.01	ND	ND	0.00	0.00	ND	
N10	F	32	N	0.05	ND	ND	0.00	0.02	ND	
N11	F	53	N	0.01	ND	ND	0.02	ND	ND	IT-
N12	F	40	N	0.01	ND	ND	0.01	ND	ND	IT-

Mugwort and component Art v 1 and Art v 3 (*A. vulgaris*), Art an 7 (*A. annua*), nAna c 2 is pineapple CCD marker bromelain, pool: Sera pool CN. IB: used for individual immunoblot, IT: intradermal test, +: positive to nArt an 7, -: negative to nArt an 7. ND: not determined.

Symptom abbreviations: AS, asthma; C, conjunctivitis; R, rhinitis; E, Eczema; N, none.

*These 10 patients appeared in a Online Repository Table E1 linked to previous publication Gao et al. JACI, 2013;131(1):223–4.

Table 2. Dutch mugwort-allergic sera ImmunoCAP specific IgE values to mugwort (*A. vulgaris*), nArt an 7 and CCD marker nAna c 2 (kU_A/L)

Sera code		sIgE (kU _A /L)				Pool
		mugwort	nArt an 7	rArt an 7	nAna c 2	
99-767604	D1	11.80	1.86	0.41	5.87	CCD+
99-774125	D2	34.80	8.30	0.48	70.59	CCD+
03-768987	D3	33.10	37.90	0.08	0.49	CCD+
04-766396	D4	31.80	1.10	0.34	1.22	CCD+
99-767616	D5	44.10	1.22	0.06	0.02	CCD-
99-772321	D6	29.50	2.80	1.54	0.08	CCD-
99-774548	D7	27.80	8.99	1.80	0.20	CCD-
01-769980	D8	39.70	2.08	0.06	0.05	CCD-
02-765603	D9	50.20	1.89	0.05	0.05	CCD-
02-768285	D10	37.10	3.91	0.05	0.03	CCD-
04-766200	D11	26.60	3.24	0.06	0.05	CCD-
04-766401	D12	44.70	2.25	0.11	0.14	CCD-
04-768067	D13	16.30	8.42	0.58	0.07	CCD-
99-766413	D14	10.90	0.74	1.06	1.02	
99-773234	D15	27.50	0.92	0.38	22.26	
00-771886	D16	14.50	0.85	0.40	4.11	
01-766764	D17	15.90	0.39	0.40	0.33	
01-770265	D18	13.80	0.58	0.57	1.75	
01-770910	D19	26.20	0.24	0.00	0.00	
04-768473	D20	36.00	2.73	0.02	0.00	

CCD+: positive to nAna c 2, -: negative to nAna c 2.

Table 3. In vivo assays of *A. annua* pollen extract and nArt an 7

Patients	Wheal diameter(mm)					IgE (kUA/L)		
	A. annua extract		nArt an 7		Histamine	mugwort	nArt an 7	
	(0.5 µg/ml)		(0.5µg/ml)		(10 µg/ml)			
P54	+	10x9	+	13x9	+	10x10	56.10	5.70
P101	+	10x10	-	0	+	12x10	1.29	0.03
P102	+	21x17	+	7x6	+	9x9	94.8	10.50
P103	+	12x10	-	0	+	10x9	5.63	0.40
P104	+	14x14	+	8x8	+	10x10	79.50	2.99
P105	+	13x13	+	6x6	+	10x10	35.70	1.00
P106	+	17x15	-	5x4	+	10x9	14.00	0.68
P107	+	7x6	-	0	+	10x10	0.49	0.34
P108	-	5x3	-	2x2	+	10x8	2.56	0.08
P109	+	7x6	+	5x5	+	7x6	45.50	0.95
P110	+	10x8	+	6x5	+	8x7	14.30	0.23
P111	+	10x8	-	2x2	+	11x10	14.60	0.39
P112	+	7x6	-	2x2	+	11x9	3.49	0.18
P113	+	9x8	+	6x5	+	10x10	16.90	0.59
N11	-	0	-	0	+	12x12	0.01	0.02
N12	-	0	-	0	+	8x8	0.01	0.01

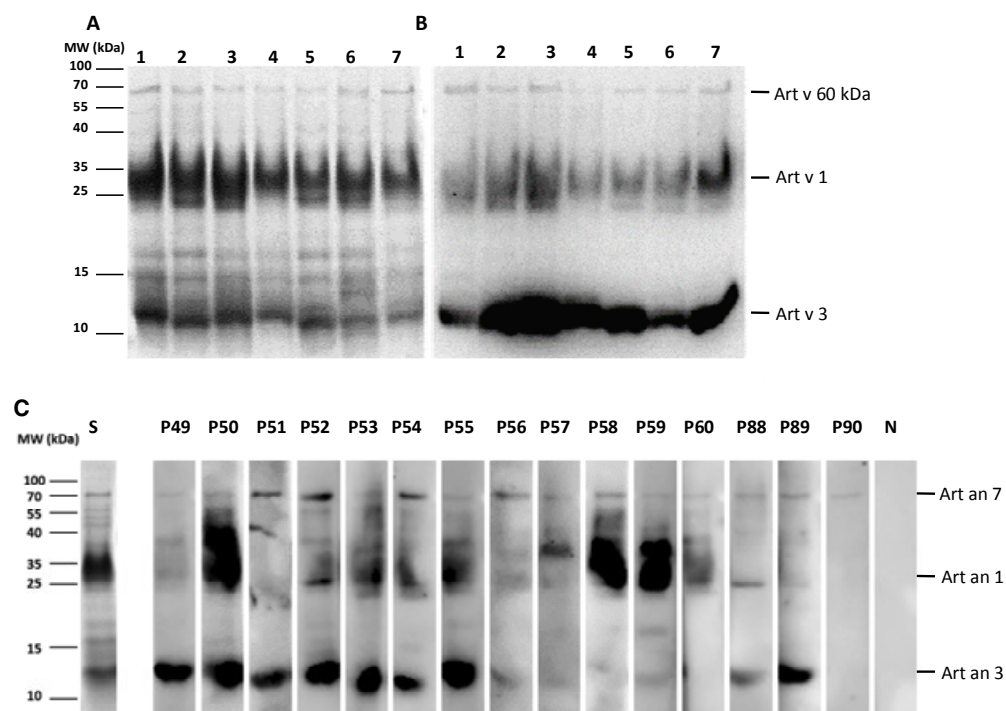


Figure 1

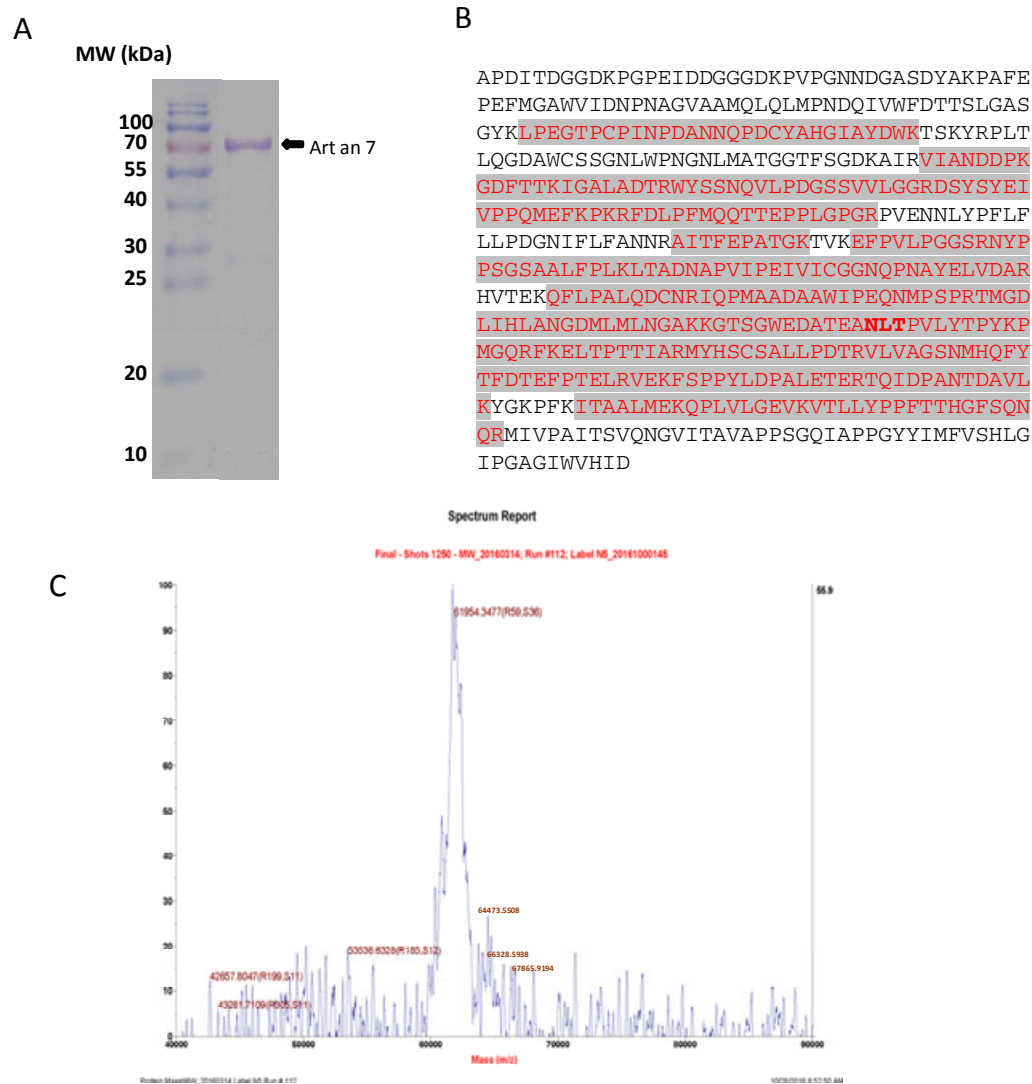


Figure 2

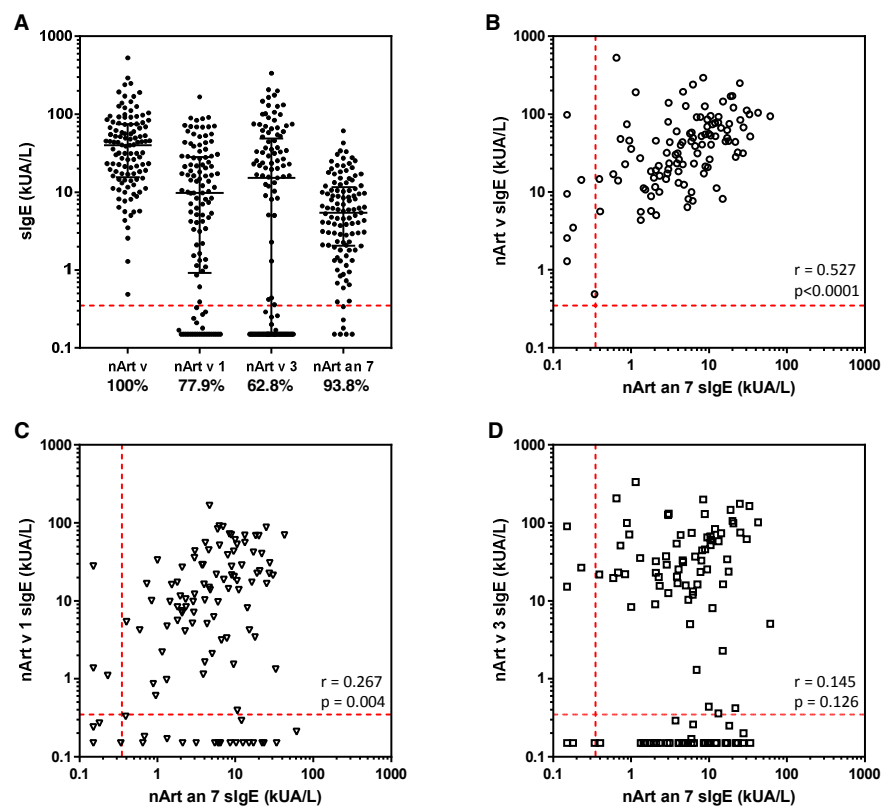


Figure 3

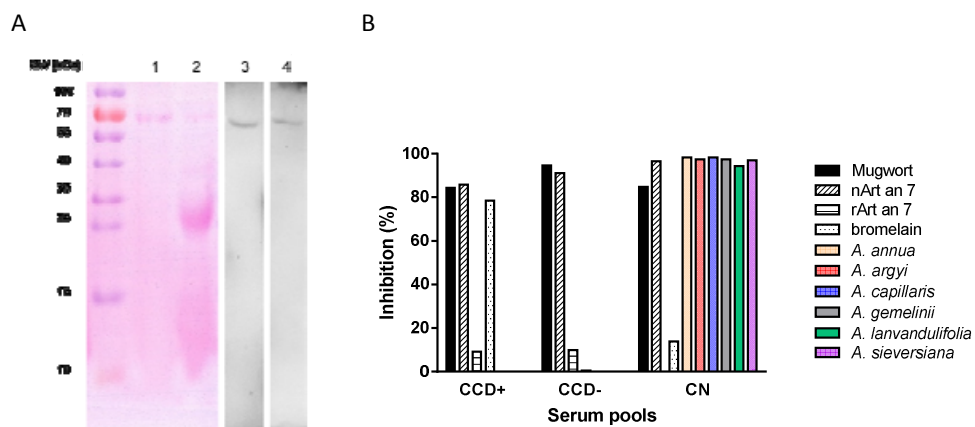
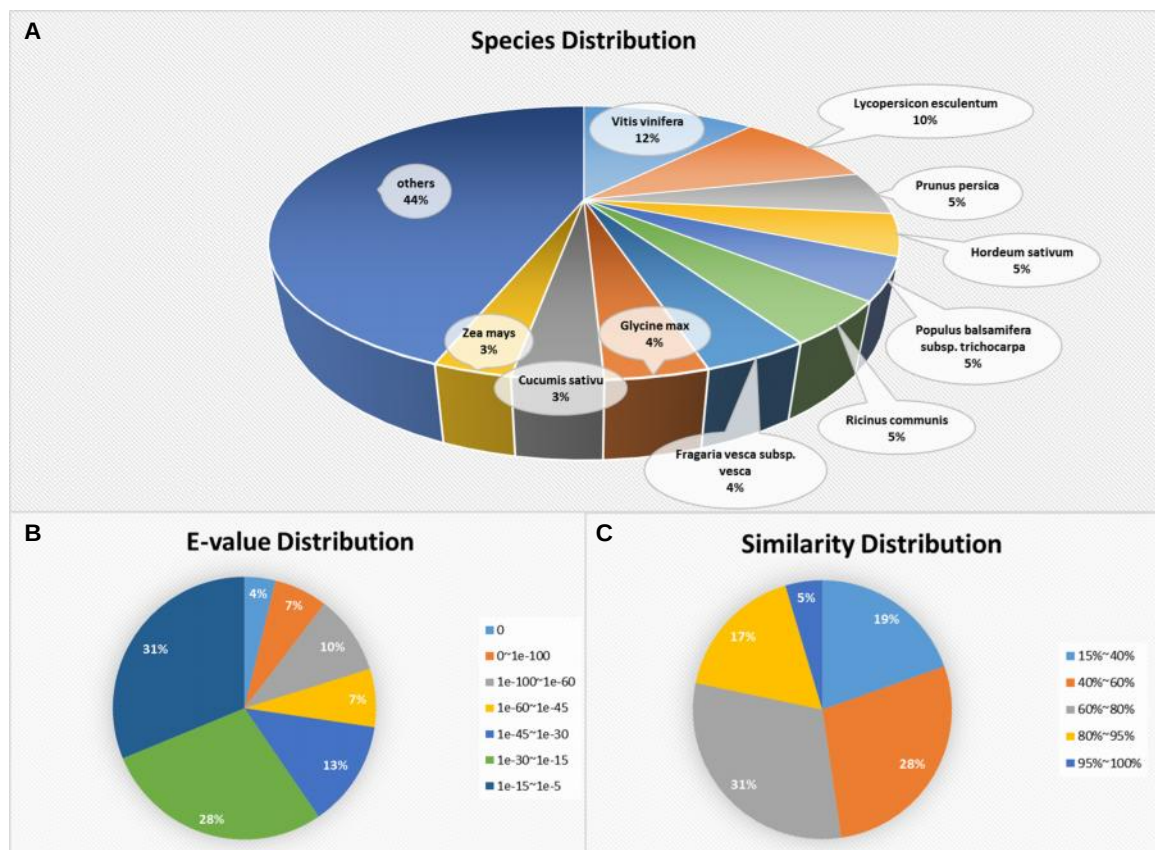
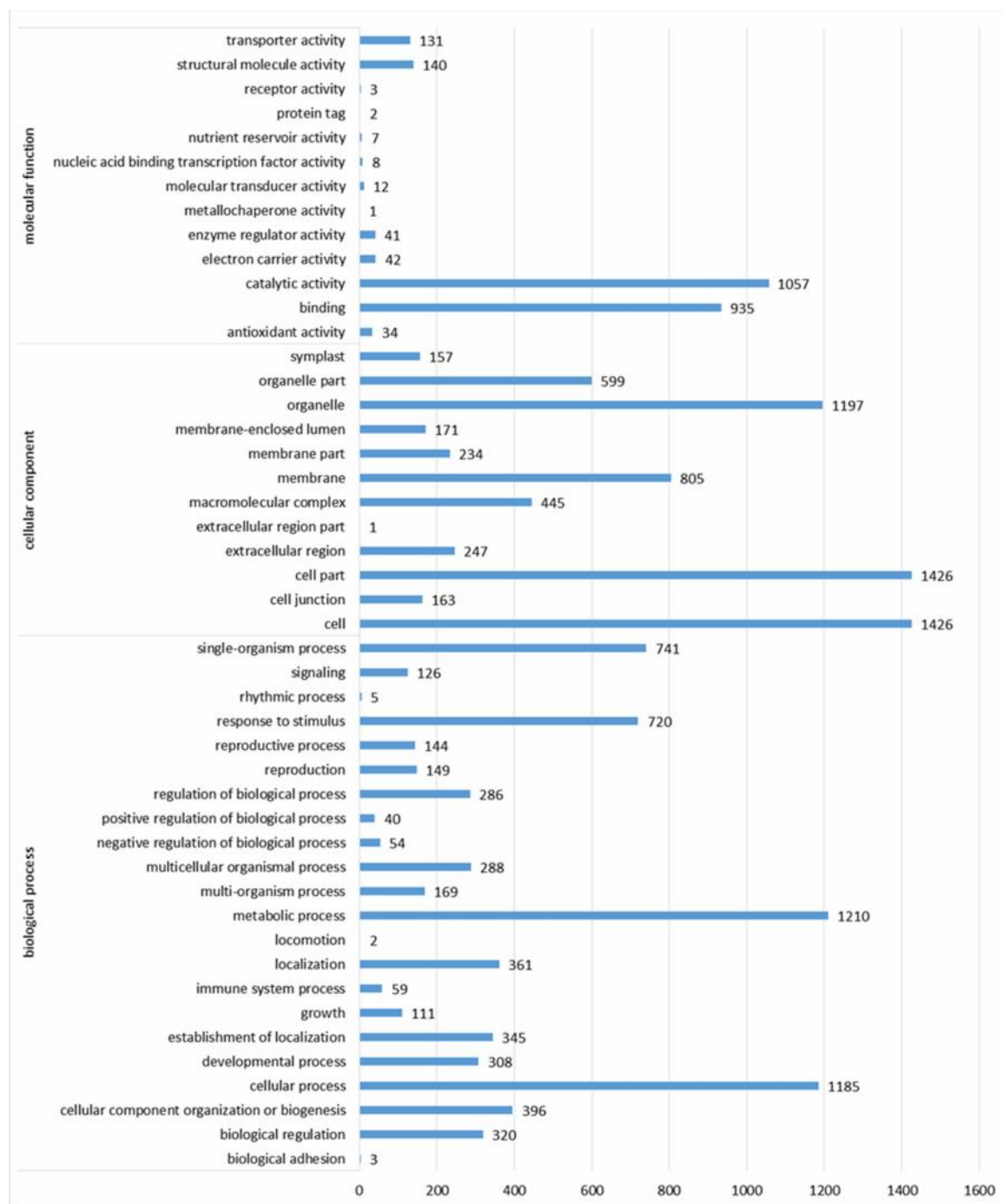


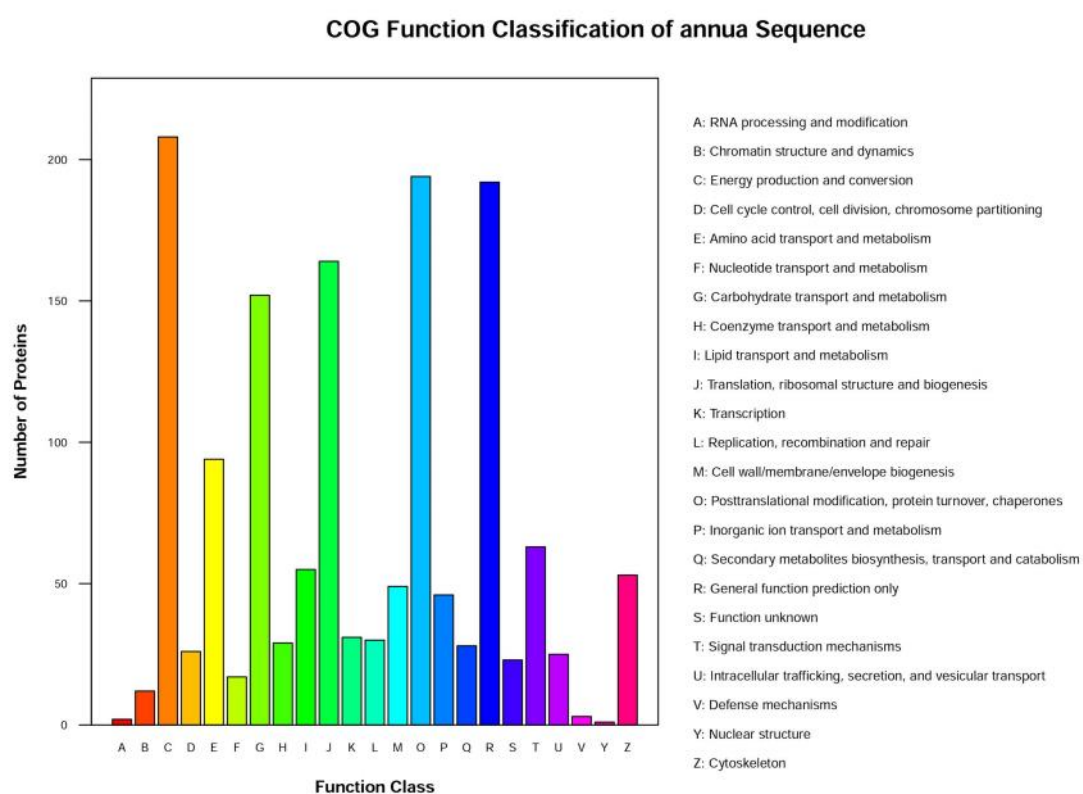
Figure 4



Appendix 1: Figure S1. NR classification of *A. annua* pollen transcriptome. The species distribution of the result of NR annotation (A) E-value distribution of the result of NR annotation (B), and the similarity distribution of the result of NR annotation (C).



Appendix 2: Figure S2. GO classification analysis of proteins in proteome. On the left are GO functions divided into three ontologies: molecular function, cellular component and biological process. The number of proteins which have the GO function is shown on the right.



Appendix 3: Figure S3 COG function classification of proteins in proteome.

The horizontal ordinate letters on the right stand for function classes of COG, and the vertical coordinates are numbers of proteins of each class.

Signal peptide

```

Art an 7.0101 : MASSIKTVILFLLPLLAYSIAAPDITDG GDKPGGEIDDDGGGDKPVPGNNDGASDYAKFAE EEFMGWVIDNPNAGVAAMQLQMFNDQIVWFDTTS : 99
Art ar 7.0101 : MASSIKTVILFLLPLLAYSIAAPDITDG GDKPGGEIDDDGGGDKPVPGNNDGASDYAKFAE EEFMGWVIDNPNAGVAAMQLQMFNDQIVWFDTTS : 99
Art ca 7.0101 : MASSIKTVILFLLPLLAYSIAAPDITDG GDKPGGEIDDDGGGDKPVPGNNDGASDYAKFAE EEFMGWVIDNPNAGVAAMQLQMFNDQIVWFDTTS : 99
Art gm 7.0101 : MASSIKTVILFLLPLLAYSIAAPDITDG GDKPGGEIDDDGGGDKPVPGNNDGASDYAKFAE EEFMGWVIDNPNAGVAAMQLQMFNDQIVWFDTTS : 99
Art la 7.0101 : MASSIKTVILFLLPLLAYSIAAPDITDG GDKPGGEIDDDGGGDKPVPGNNDGASDYAKFAE EEFMGWVIDNPNAGVAAMQLQMFNDQIVWFDTTS : 99
Art si 7.0101 : MASSIKTVILFLLPLLAYSIAAPDITDG GDKPGGEIDDDGGGDKPVPGNNDGASDYAKFAE EEFMGWVIDNPNAGVAAMQLQMFNDQIVWFDTTS : 100
Art v 7.0101 : MASSIKTVILFLLPLLAYSIAAPDITDG GDKPGGEIDDDGGGDKPVPGNNDGASDYAKFAE EEFMGWVIDNPNAGVAAMQLQMFNDQIVWFDTTS : 99

Art an 7.0101 : LGSGYKLPSTPCNNPLANNQPDCTVHGIAIDWRTSKYRPLTLQGLAWCSSGNLWPNGMNATGTFSGDRAIRVINDNDNGDFTTKIGALADTRWY : 199
Art ar 7.0101 : LGSGYKLPSTPCNNPLANNQPDCTVHGIAIDWRTSKYRPLTLQGLAWCSSGNLWPNGMNATGTFSGDRAIRVINDNDNGDFTTKIGALADTRWY : 199
Art ca 7.0101 : LGSGYKLPSTPCNNPLANNQPDCTVHGIAIDWRTSKYRPLTLQGLAWCSSGNLWPNGMNATGTFSGDRAIRVINDNDNGDFTTKIGALADTRWY : 199
Art gm 7.0101 : LGSGYKLPSTPCNNPLANNQPDCTVHGIAIDWRTSKYRPLTLQGLAWCSSGNLWPNGMNATGTFSGDRAIRVINDNDNGDFTTKIGALADTRWY : 199
Art la 7.0101 : LGSGYKLPSTPCNNPLANNQPDCTVHGIAIDWRTSKYRPLTLQGLAWCSSGNLWPNGMNATGTFSGDRAIRVINDNDNGDFTTKIGALADTRWY : 199
Art si 7.0101 : LGSGYKLPSTPCNNPLANNQPDCTVHGIAIDWRTSKYRPLTLQGLAWCSSGNLWPNGMNATGTFSGDRAIRVINDNDNGDFTTKIGALADTRWY : 200
Art v 7.0101 : LGSGYKLPSTPCNNPLANNQPDCTVHGIAIDWRTSKYRPLTLQGLAWCSSGNLWPNGMNATGTFSGDRAIRVINDNDNGDFTTKIGALADTRWY : 199

Art an 7.0101 : SSNQVLPDGSSVVLGGRDSYSYEVPPQMEFKRRFDLPQQQTTEPPLGGRPVENNLYPFFILLPDGNIFFLANNRAITFEFATGKTVEKFPVLPGGS : 299
Art ar 7.0101 : SSNQVLPDGSSVVLGGRDSYSYEVPPQMEFKRRFDLPQQQTTEPPLGGRPVENNLYPFFILLPDGNIFFLANNRAITFEFATGKTVEKFPVLPGGS : 299
Art ca 7.0101 : SSNQVLPDGSSVVLGGRDSYSYEVPPQMEFKRRFDLPQQQTTEPPLGGRPVENNLYPFFILLPDGNIFFLANNRAITFEFATGKTVEKFPVLPGGS : 299
Art gm 7.0101 : SSNQVLPDGSSVVLGGRDSYSYEVPPQMEFKRRFDLPQQQTTEPPLGGRPVENNLYPFFILLPDGNIFFLANNRAITFEFATGKTVEKFPVLPGGS : 299
Art la 7.0101 : SSNQVLPDGSSVVLGGRDSYSYEVPPQMEFKRRFDLPQQQTTEPPLGGRPVENNLYPFFILLPDGNIFFLANNRAITFEFATGKTVEKFPVLPGGS : 299
Art si 7.0101 : SSNQVLPDGSSVVLGGRDSYSYEVPPQMEFKRRFDLPQQQTTEPPLGGRPVENNLYPFFILLPDGNIFFLANNRAITFEFATGKTVEKFPVLPGGS : 300
Art v 7.0101 : SSNQVLPDGSSVVLGGRDSYSYEVPPQMEFKRRFDLPQQQTTEPPLGGRPVENNLYPFFILLPDGNIFFLANNRAITFEFATGKTVEKFPVLPGGS : 299

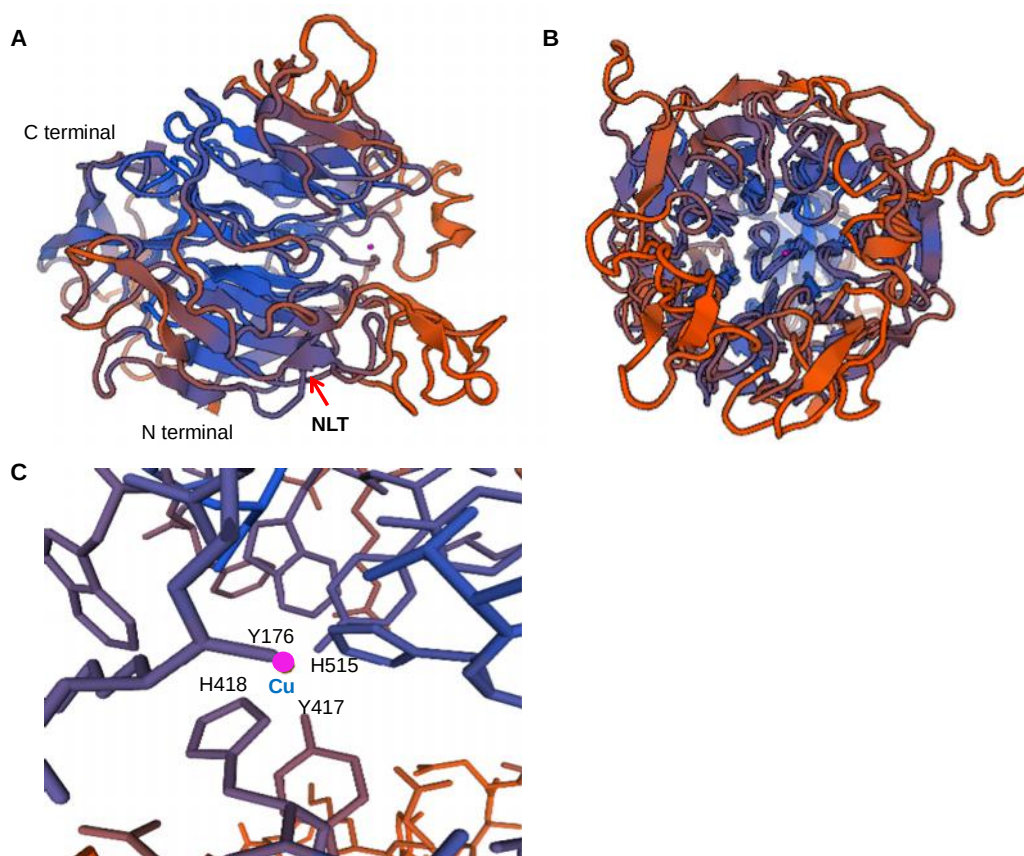
Art an 7.0101 : RNYPSSGSAALFPLKLTADNAPVPEIVICGGNENAYELVTARHVTKEFLFALQDCNRIDFMAALAAWIPEQNMPSPTMTGDIHIANGCMMLMGAV : 399
Art ar 7.0101 : RNYPSSGSAALFPLKLTADNAPVPEIVICGGNENAYELVTARHVTKEFLFALQDCNRIDFMAALAAWIPEQNMPSPTMTGDIHIANGCMMLMGAV : 399
Art ca 7.0101 : RNYPSSGSAALFPLKLTADNAPVPEIVICGGNENAYELVTARHVTKEFLFALQDCNRIDFMAALAAWIPEQNMPSPTMTGDIHIANGCMMLMGAV : 399
Art gm 7.0101 : RNYPSSGSAALFPLKLTADNAPVPEIVICGGNENAYELVTARHVTKEFLFALQDCNRIDFMAALAAWIPEQNMPSPTMTGDIHIANGCMMLMGAV : 399
Art la 7.0101 : RNYPSSGSAALFPLKLTADNAPVPEIVICGGNENAYELVTARHVTKEFLFALQDCNRIDFMAALAAWIPEQNMPSPTMTGDIHIANGCMMLMGAV : 399
Art si 7.0101 : RNYPSSGSAALFPLKLTADNAPVPEIVICGGNENAYELVTARHVTKEFLFALQDCNRIDFMAALAAWIPEQNMPSPTMTGDIHIANGCMMLMGAV : 400
Art v 7.0101 : RNYPSSGSAALFPLKLTADNAPVPEIVICGGNENAYELVTARHVTKEFLFALQDCNRIDFMAALAAWIPEQNMPSPTMTGDIHIANGCMMLMGAV : 399

Art an 7.0101 : RGTSGWELATE NLTFLVLYTPYKFMGQRFKELEPTTIARMYHSCSALLPDTRVLVAGSNMHQFYTFDTEFPTELRLVEKFSPPYLDFALEHERRGIDFANT : 499
Art ar 7.0101 : RGTSGWELATE NLTFLVLYTPYKFMGQRFKELEPTTIARMYHSCSALLPDTRVLVAGSNMHQFYTFDTEFPTELRLVEKFSPPYLDFALEHERRGIDFANT : 499
Art ca 7.0101 : RGTSGWELATE NLTFLVLYTPYKFMGQRFKELEPTTIARMYHSCSALLPDTRVLVAGSNMHQFYTFDTEFPTELRLVEKFSPPYLDFALEHERRGIDFANT : 499
Art gm 7.0101 : RGTSGWELATE NLTFLVLYTPYKFMGQRFKELEPTTIARMYHSCSALLPDTRVLVAGSNMHQFYTFDTEFPTELRLVEKFSPPYLDFALEHERRGIDFANT : 499
Art la 7.0101 : RGTSGWELATE NLTFLVLYTPYKFMGQRFKELEPTTIARMYHSCSALLPDTRVLVAGSNMHQFYTFDTEFPTELRLVEKFSPPYLDFALEHERRGIDFANT : 499
Art si 7.0101 : RGTSGWELATE NLTFLVLYTPYKFMGQRFKELEPTTIARMYHSCSALLPDTRVLVAGSNMHQFYTFDTEFPTELRLVEKFSPPYLDFALEHERRGIDFANT : 500
Art v 7.0101 : RGTSGWELATE NLTFLVLYTPYKFMGQRFKELEPTTIARMYHSCSALLPDTRVLVAGSNMHQFYTFDTEFPTELRLVEKFSPPYLDFALEHERRGIDFANT : 499

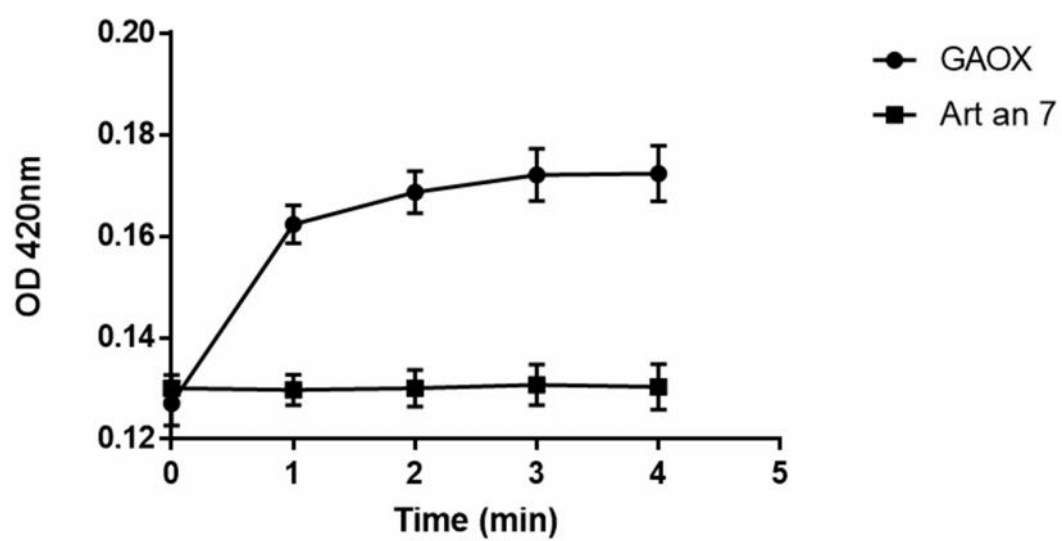
Art an 7.0101 : DVVLKYGKPKITAAIMEKQPLVLGEVKVTLLYPFTTHGFSQNCRMIVFAITVWNGVITAAPPSSGIAPPGYIIMFVSHLGIPGAGIWWHII : 594
Art ar 7.0101 : DVVLKYGKPKITAAIMEKQPLVLGEVKVTLLYPFTTHGFSQNCRMIVFAITVWNGVITAAPPSSGIAPPGYIIMFVSHLGIPGAGIWWHII : 594
Art ca 7.0101 : DVVLKYGKPKITAAIMEKQPLVLGEVKVTLLYPFTTHGFSQNCRMIVFAITVWNGVITAAPPSSGIAPPGYIIMFVSHLGIPGAGIWWHII : 594
Art gm 7.0101 : DVVLKYGKPKITAAIMEKQPLVLGEVKVTLLYPFTTHGFSQNCRMIVFAITVWNGVITAAPPSSGIAPPGYIIMFVSHLGIPGAGIWWHII : 594
Art la 7.0101 : DVVLKYGKPKITAAIMEKQPLVLGEVKVTLLYPFTTHGFSQNCRMIVFAITVWNGVITAAPPSSGIAPPGYIIMFVSHLGIPGAGIWWHII : 594
Art si 7.0101 : DVVLKYGKPKITAAIMEKQPLVLGEVKVTLLYPFTTHGFSQNCRMIVFAITVWNGVITAAPPSSGIAPPGYIIMFVSHLGIPGAGIWWHII : 595
Art v 7.0101 : DVVLKYGKPKITAAIMEKQPLVLGEVKVTLLYPFTTHGFSQNCRMIVFAITVWNGVITAAPPSSGIAPPGYIIMFVSHLGIPGAGIWWHII : 594

```

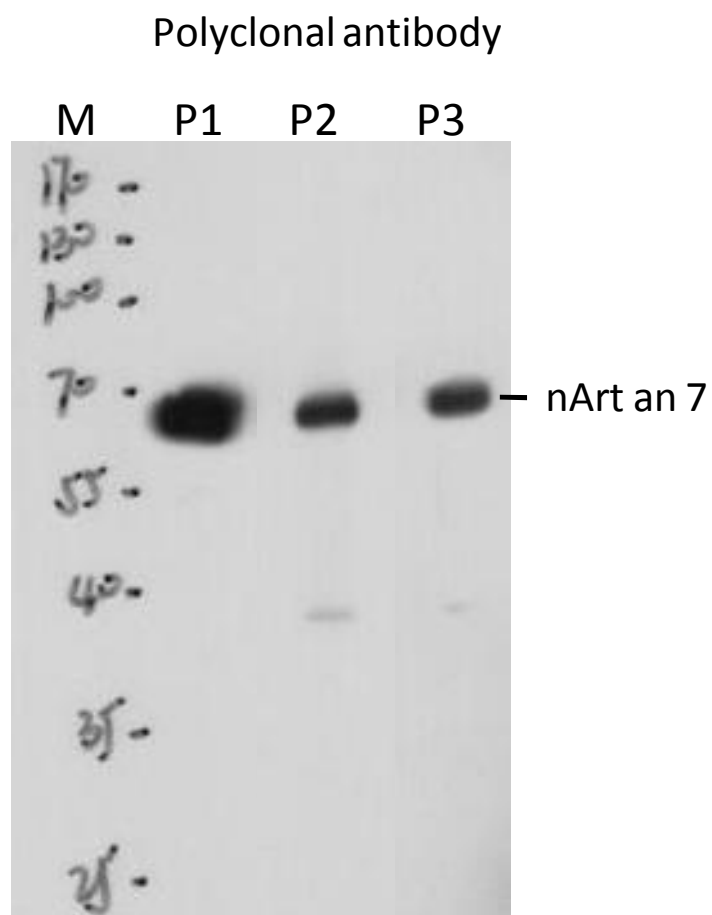
Appendix 5: Figure S4. Deduced galactose oxidase precursor protein sequences in seven *Artemisa pollens* (NCBI: KY428925-KY428931). Glycosylation site was boxed.



Appendix 6 Figure S5 3D model of Art an 7 protein with Blast and HHBlits against the SWISS-MODEL template 4unm. a, b: Typical kelch beta-propeller structure 3D views from different angle, pink dot is Cu, N-glycosylation site NLT locates on the surface. c: the active site showing ligand of copper and residues.



Appendix 7 Figure S6 Enzymatic activity on D-galactose of nArt an 7 and GAOX from *Dactylium dendroides*



Appendix 8 Figure S7 Reaction of polyclonal antibodies derived from three synthesized peptides from Art an 7 amino acid sequence to naturally purified Art an 7.

Rabbit polyclonal antibody (concentration 0.3mg/ml) dilution

for p1, 1:5000 ; p2 and p3, 1:500

Peptide sequence and position in Art an 7 mature protein

P1 aa 1-14, APDITDGGDKPGPE

P2 aa 508-520, LYPPFTTHGFSQN

P3 aa 377-389, KGTSGWEDATEANL

```

      *      20      *      40      *      60      *      80      *      100
Art v 1.0101 : AGSKLCEKTSKTVSGKCDNKKCKKCIWEHFAQHGACHKREAGKESCFYFDCSKSPPGATTAPPGAAPPPAAGGSRSPFADGGSPPPFADGGSPFVDDG : 100
Art an 1.0101 : AGSKLCEKTSKTVSGKCDNKKCKKCIWEHFAQHGACHKREAGKESCFYFDCSKSPPGATTAPPGAAPPPAAGGSRSPFADGGSPPPFADGGSPFVDDG : 100

      *      20      *      40      *      60      *      80      *      100
Art v 1.0101 : SPPPPSP : 108
Art an 1.0101 : SPPPPSP : 108

      *      20      *      40      *      60      *      80      *      100
Art v 2.0101 : HETYGEPGNTPDYVHAHNCIRRVLCMKPLCWLSTIGKVAQAAAEITRPDCSLVBSDRCGENMAQGAINGSMVQLWDERLDYDYNENKCIKMGHYTC : 99
Art an 2.0101 : HETYGEPGNTPDYVHAHNCIRRVLCMKPLCWLSTIGKVAQAAAEITRPDCSLVBSDRCGENMAQGAINGSMVQLWDERLDYDYNENKCIKMGHYTC : 100

      *      120      *      140
Art v 2.0101 : IVKANSERVGGGRCALCSNGKAYIVVCNYDPPGNVVGQKPY : 139
Art an 2.0101 : IVKANSERVGGGRCALCSNGKAYIVVCNYDPPGNVVGQKPY : 140

      *      20      *      40      *      60      *      80      *      100
Art v 3.0201 : TLLGSDVANKIDTCNLHAGGGEVFDGCTGVKGLNCAKTTTDDROACNCLATFESKNDLPSDAVLPSCQNNIPYKSLETDCNKVK* : 91
Art v 3.0202 : TLLGSDVANKIDTCNLHAGGGEVFDGCTGVKGLNCAKTTTDDROACNCLATFESKNDLPSDAVLPSCQNNIPYKSLETDCNKVK* : 91
Art v 3.0301 : TLLGSDVANKIDTCNLHAGGGEVFDGCTGVKGLNCAKTTTDDROACNCLATFESKNDLPSDAVLPSCQNNIPYKSLETDCNKVK* : 92
Art an 3.0401 : TLLGSDVANKIDTCNLHAGGGEVFDGCTGVKGLNCAKTTTDDROACNCLATFESKNNRRENAIVLPSCQNNIPYKSLETDCNKVK* : 92

      *      20      *      40      *      60      *      80      *      100
Art v 4.0101 : MSWQTYVDDHLMCDIEGTGQHLTSAATFGTDGTVAKSASFPEFKPNEILAIKEFNEAGQIAPTGLFLGCAKYMVLQGEACAVIRGKKGAGGICIKKTC : 100
Art an 4.0101 : MSWQTYVDDHLMCDIEGTGQHLTSAATFGTDGTVAKSASFPEFKPNEILAIKEFNEAGQIAPTGLFLGCAKYMVLQGEACAVIRGKKGAGGICIKKTC : 100

      *      120
Art v 4.0101 : CAMVFGIYDEPVAPGQCNMVERLGDYLLDQNM : 133
Art an 4.0101 : CAMVFGIYDEPVAPGQCNMVERLGDYLLDQNM : 133

      *      20      *      40      *      60      *      80      *      100
Art v 4.0201 : MSWQTYVDDHLMCDIEGTGQHLTAAAILGLDGTVAKSDFPEFKPEEMKGIINEFNEVGTIAPTGLFLGCAKYMVLQGEACAVIRGKKGAGGICIKKTC : 100
Art an 4.0201 : MSWQTYVDDHLMCDIEGTGQHLTAAAILGLDGTVAKSDFPEFKPEEMKGIINEFNEVGTIAPTGLFLGCAKYMVLQGEACAVIRGKKGAGGICIKKTC : 100

      *      120
Art v 4.0201 : CAMVMGIYDEPVAPGQCNMIVERLGDYLLDQNM : 133
Art an 4.0201 : CAMVMGIYDEPVAPGQCNMIVERLGDYLLDQNM : 133

      *      20      *      40      *      60      *      80
Art v 5.0101 : MADEKRAECDRIFGAFDRNGDGKISAAELGESIKLGSVSPPEEVQTMMDLDTDDGDYISYDEFAEFFNANRGLMKDVGKIF : 82
Art an 5.0101 : MADEKRAECDRIFGAFDRNGDGKISAAELGESIKLGSVSPPEEVQTMMDLDTDDGDYISYDEFAEFFNANRGLMKDVGKIF : 82

      *      20      *      40      *      60      *      80      *      100
Art v 6.0101 : RADIGDELEAACNSTRRLGHECAANIIDKWRCKADWEKNRCAIAKCAQGFAGKTTGGLGGEIYVVTDCSDNAANPKPGTLRGVTDQKPLWIIIFKN : 100
Art an 6.0101 : RADIGDELEAACNSTRRLGHECAANIIDKWRCKADWEKNRCAIAKCAQGFAGKTTGGLGGEIYVVTDCSDNAANPKPGTLRGVTDQKPLWIIIFKN : 99

      *      120      *      140      *      160      *      180      *      200
Art v 6.0101 : CMVILKHELIVINKDKTIDGRGANVEITCGGLTIHNVCNVIHNIHIDIKVTEGGIIKATTAKPGHRRHKSDDGICVAGSSKIWIHCTLSHGPDGLIH : 200
Art an 6.0101 : CMVILKHELIVINKDKTIDGRGANVEITCGGLTIHNVCNVIHNIHIDIKVTEGGIIKATTAKPGHRRHKSDDGICVAGSSKIWIHCTLSHGPDGLIH : 199

      *      220      *      240      *      260      *      280      *      300
Art v 6.0101 : VTILGSTAVTISNCRFSHHQKIDLLGALNSHVDDKKMHVTVAFNRFARACDQFMFRCRFGFFQVNNNDYTSWGTYAIGGSANPTILSQGNRFHAPNEMKN : 300
Art an 6.0101 : VTILGSTAVTISNCRFSHHQKIDLLGALNSHVDDKKMHVTVAFNRFARACDQFMFRCRFGFFQVNNNDYTSWGTYAIGGSANPTILSQGNRFHAPNEMKN : 299

      *      320      *      340      *      360      *
Art v 6.0101 : NVLVFALAPFESMKWNNRSEKDLLENCAIFVASGCDPHLTPEQKSHLIFAEPGSAVLQLTSCAGTLKCVPGKPG : 375
Art an 6.0101 : NVLVFALAPFESMKWNNRSEKDLLENCAIFVASGCDPHLTPEQKSHLIFAEPGSAVLQLTSCAGTLKCVPGKPG : 374

```

Appendix 10: Figure S8. Deduced mature peptides sequences in *A. annua* proteome of seven Art v allergens (*A. vulgaris*) homologous.

Appendix 4: Table S1. Top 10 identified proteins in the *Artemisia annua* pollen

Protein ID	Annotation	Protein Mass	Coverage	pI	# unique peptide	# unique spectrum	# total spectrum
CL1968.Contig1	polygalacturonase	42661.53	67.77	8.59	26	1156	1317
Unigene154	NtPRp27-like partial	22132.11	81.68	8.35	20	1224	1383
Unigene5168	esterase/lipase 6	34810.44	52.3	9.22	14	867	918
Unigene20457	reticuline oxidase-like protein-like	56527.22	65.54	9.59	29	1196	1363
CL2164.Contig7	hypothetical protein	16893.91	93.96	3.88	6	806	979
CL2335.Contig4	Amb a 1-like protein	40273.21	80.06	6.08	23	719	783
CL530.Contig1	fructose-bisphosphate aldolase	38830	73.11	7.39	14	439	627
Unigene14150	subtilisin-like protease-like	71801.93	45.35	7.49	24	656	733
Unigene18656	Putative galactose oxidase [<i>Helianthus annuus</i>] OTG34910	62067.55	64.37	4.75	20	709	792
CL1179.Contig1	enolase-like	48150.52	71.4	5.75	17	364	456
CL530.Contig1	fructose-bisphosphate aldolase	38830	73.11	7.39	14	439	627
Unigene14150	subtilisin-like protease-like	71801.93	45.35	7.49	24	656	733
Unigene18656	Putative galactose oxidase [<i>Helianthus annuus</i>] OTG34910	62067.55	64.37	4.75	20	709	792
CL1179.Contig1	enolase-like	48150.52	71.4	5.75	17	364	456

Appendix 9: Table S1. Art an 7 homologous sequences in vegetables, fruits and spices with high identity.

Species	Identity (%)	Positives (%)	E-value	Score	Accessions
<i>Helianthus annuus</i>	74	84	0	917	OTG34910.1
<i>Cynara cardunculus</i>	73	85	0	850	gi 976928673 gb KVI11588.1
<i>Coffea canephora</i>	52	68	1.2E-180	542	gi 661893486 emb CDP03295.1
<i>Daucus carota</i>	50	65	1.1E-177	524	gi 1040920028 ref XP_017222277.1 ;gi 1021026111 gb KZM83898.1
<i>Solanum lycopersicum</i>	49	67	1.1E-177	523	gi 723714187 ref XP_004242940.2
<i>Solanum tuberosum</i>	49	67	2.5E-177	521	gi 565348545 ref XP_006341271.1
<i>Solanum pennellii</i>	49	67	3.5E-175	519	gi 970039462 ref XP_015081068.1
<i>Capsicum annuum</i>	49	67	8.3E-175	521	gi 1026042282 ref XP_016567344.1
<i>Sesamum indicum</i>	49	65	8.3E-170	505	gi 747063489 ref XP_011078293.1
<i>Vitis vinifera</i>	48	65	2.6E-169	504	gi 731406827 ref XP_010656292.1
<i>Cicer arietinum</i>	46	63	2.6E-165	493	gi 502164005 ref XP_004513017.1
<i>Pyrus x bretschneideri</i>	45	62	5.3E-162	486	gi 694417780 ref XP_009336933.1
<i>Malus domestica</i>	45	62	7.7E-160	481	gi 658054398 ref XP_008362947.1
<i>Phaseolus vulgaris</i>	45	62	1.9E-158	476	gi 593705035 ref XP_007152887.1 ;gi 561026196 gb ESW24881.1
<i>Prunus mume</i>	43	62	1.3E-156	474	gi 645274911 ref XP_008242563.1
<i>Ziziphus jujuba</i>	47	62	3.6E-156	469	gi 1009160647 ref XP_015898463.1
<i>Citrus sinensis</i>	44	60	2.4E-155	467	gi 568843607 ref XP_006475693.1
<i>Prunus persica</i>	43	62	4.7E-155	471	gi 595810622 ref XP_007203106.1 ;gi 462398637 gb EMJ04305.1
<i>Cucumis melo</i>	44	60	1.7E-154	466	gi 659071612 ref XP_008460895.1
<i>Cucumis sativus</i>	46	62	2E-154	466	gi 449443335 ref XP_004139435.1 ;gi 700205483 gb KGN60602.1
<i>Morus notabilis</i>	45	62	5.1E-153	462	gi 703140955 ref XP_010107380.1 ;gi 587928679 gb EXC15869.1
<i>Glycine max</i>	43	61	2.3E-152	460	gi 356495218 ref XP_003516476.1 ;gi 734352493 gb KHN13139.1 ;gi 947128465 gb KRN76319.1
<i>Brassica rapa</i>	45	63	7.9E-151	456	gi 685272978 ref XP_009127543.1
<i>Raphanus sativus</i>	44	61	8.1E-151	455	gi 1072989268 ref XP_018445339.1