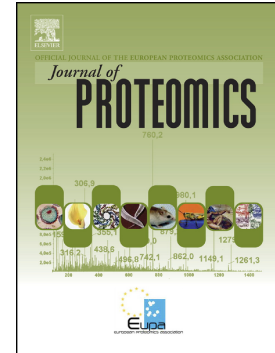


Accepted Manuscript

Charting novel allergens from date palm pollen (*Phoenix sylvestris*) using homology driven proteomics

Bodhisattwa Saha, Naren Pandey, Swati Gupta Bhattacharya



PII: S1874-3919(17)30188-4
DOI: doi: [10.1016/j.jprot.2017.05.021](https://doi.org/10.1016/j.jprot.2017.05.021)
Reference: JPROT 2860

To appear in: *Journal of Proteomics*

Received date: 1 December 2016
Revised date: 5 May 2017
Accepted date: 17 May 2017

Please cite this article as: Bodhisattwa Saha, Naren Pandey, Swati Gupta Bhattacharya, Charting novel allergens from date palm pollen (*Phoenix sylvestris*) using homology driven proteomics, *Journal of Proteomics* (2017), doi: [10.1016/j.jprot.2017.05.021](https://doi.org/10.1016/j.jprot.2017.05.021)

This is a PDF file of an unedited manuscript that has been accepted for publication. As a service to our customers we are providing this early version of the manuscript. The manuscript will undergo copyediting, typesetting, and review of the resulting proof before it is published in its final form. Please note that during the production process errors may be discovered which could affect the content, and all legal disclaimers that apply to the journal pertain.

Charting novel allergens from Date Palm pollen (*Phoenix sylvestris*) using Homology driven proteomics

Bodhisattwa Saha¹, Naren pandey², Swati Gupta Bhattacharya^{1*}

¹ Division of Plant Biology, Bose Institute, 93/1 Acharya Prafulla Chandra Road, Kolkata 700009.

² Department of Allergy and Asthma, Belle View Clinic, 9, Dr. U.N. Brahmachari Street, Kolkata 700017.

*Corresponding author

Address of Correspondence:

Prof Swati Gupta Bhattacharya, Division of Plant Biology, Bose Institute, 93/1 Acharya Prafulla Chandra Road, Kolkata 700009. Email: swati@jcbose.ac.in. Phone Number. +91-33-23031131. Fax: +91-33 23506790.

List of abbreviations:

MALDI-TOF: Matrix Assisted Laser Desorption Ionisation Time of Flight, IgE: Immunoglobulin E, SPT: Skin Prick Test, pNPP: *p*-Nitrophenyl Phosphate, SNP: Single Nucleotide Polymorphism.

Keywords

Date Palm, *de novo* sequencing, Pollen, allergen, Cross-reactivity, proteomics.

Highlights

- Date palm pollen is known to be a major aeroallergen in India.
- Ten allergens identified from date palm pollen for the first time.
- *De novo* sequencing and validation by acetylation used to identify the allergens.
- Glycan or Peptidic IgE binding epitopes diagnosed on the allergens.

Abstract:

Pollen grains from *Phoenix sylvestris* (Date palm), a commonly cultivated tree in India has been found to cause severe allergic diseases in an increasing percentage of hypersensitive individuals. To unearth its allergenic components, pollen protein were profiled by two-dimensional gel electrophoresis followed by immunoblotting with date palm pollen sensitive patient sera. Allergens were identified by MALDI-TOF/TOF employing a layered proteomic approach combining conventional database dependant search and manual *denovo* sequencing followed by homology-based search as *Phoenix sylvestris* is unsequenced. Derivatization of tryptic peptides by acetylation has been demonstrated to differentiate the 'b' from the 'y' ions facilitating efficient *de novo* sequencing. Ten allergenic proteins were identified, out of which six showed homology with known allergens while others were reported for the first time. Amongst these, isoflavone reductase, beta-conglycinin, S-adenosyl methionine synthase, 1, 4 glucan synthase and beta-galactosidase were commonly reported as allergens from coconut pollen and presumably responsible for cross-reactivity. One of the allergens had IgE binding epitope recognized by its glycan moiety. The allergenic potency of date palm pollen has been demonstrated using in vitro tests. The identified allergens can be used to develop vaccines for immunotherapy against date palm pollen allergy.

The significance of the study:

Identification of allergenic proteins from sources harboring them is essential in developing therapeutic interventions. This is the first comprehensive study on the identification of allergens from *Phoenix sylvestris* (Date palm) pollen, one of the major aeroallergens in India using a proteomic approach. Proteomic methods are being increasingly used to identify allergens. However, since many of these proteins arise from species which are un-sequenced, it becomes difficult to interpret those using conventional proteomics. Date palm being an unsequenced species, the IgE-reactive proteins have been identified using a stratified proteomic workflow incorporating manual *de novo* sequencing and homology-based proteomics. This study also gives an insight into the presence of glycan nature of the IgE binding epitopes. Five proteins have been found to be common with coconut pollen allergens and presumably responsible for cross-reactivity. These can be used in diagnostics to differentiate patient cohorts allergic to both coconut and date palm pollen from true date palm pollen allergic subjects. This would also determine better specific immunotherapy regimes between the two cohorts. The allergens identified herein have potential towards vaccine development in date palm pollen allergy as well as in enriching the existing catalogue of allergenic proteins.

1. Introduction

Allergy is an Immunoglobulin E-mediated type 1 hypersensitive disorder and is caused by specific proteins called allergens affecting more than 25% of the world population [1]. When allergens interact with the immune system, mast cells and basophils are activated, causing them to release pro-inflammatory molecules like histamine, leukotrienes, and cytokines creating the allergic phenotype. Pollen grains harbour few such proteins which constitute a major part of the aeroallergen load. Susceptible individuals exposed to such pollen grains often suffer from seasonal allergic rhinitis, dermatitis, rhinoconjunctivitis and bronchial asthma.

Date palm is a flowering tree belonging to the palm family (Arecaceae) known for its edible sweet fruit and is cultivated widely in many tropical and subtropical regions of the world. In India, the most common date palm tree is *Phoenix sylvestris* (PS) also commonly called as sugar date palm. It earns revenue for the country for fruit as well as its sap which is boiled to prepare palm jaggery, a delicacy in Eastern India. However, its pollen grains have been reported to cause allergy in an increasing number of sensitized subjects. Date palm pollen allergy is seasonal which increases during its flowering time between January to March [2]. More than 43% were found to be allergic to date palm pollen in a cohort of 540 patients surveyed during a year in Kolkata metropolis [2]. Cross-reactivity had also been reported to be present among four allergenic palm pollen grains namely *Cocos nucifera*, *Borassus flabellifer*, *Phoenix sylvestris* and *Areca catechu* suggesting presence common IgE binding proteins [3]. Date palm pollen allergy has also been reported from Mediterranean and Middle East populations [4, 5].

Identification of candidate allergens is a prerequisite for initiating vaccine development for immunotherapy. Though there are reports on the epidemiology of date palm pollen allergy,

very little effort has been undertaken to identify the allergens. Moreover, identification of cross-reactive allergens will boost development of better vaccines against a wide variety of pollen species. Only one protein has been extensively characterized and identified as profilin (Pho d 2) from *Phoenix dactylifera*, a date palm variety reactive to a Spanish cohort [6]. In PS pollen, only two major IgE reactive bands of 33kD and 66kD has been determined but none have been identified [2].

Use of proteomics has paced up the discovery of novel allergens from different allergenic pollen grains as well as other species in the last decade compared to other allergen identification techniques. However, identification of proteins using proteomics is challenging in absence of sequence information. Since many of the allergens arise from sources which are non-model organisms, they remain largely unidentified and hence efficient protein identification pipelines from such species need to be developed. In such cases, *de novo* sequencing is often employed to sequence the peptides directly from the spectra without relying on any database and searched for homology. Several algorithms have been developed to automate this task and used in the high-throughput analysis. Despite such efforts, determination of accurate sequence is far from complete and better algorithms are still in quest. Manual *de novo* sequencing is an alternative to such softwares which are often expensive and can be applied when annotating smaller datasets [7].

In this study, we have used an immuno-proteomics approach to identify allergens from PS pollen for the first time. A stratified proteomic methodology employing computer assisted manual *de novo* sequencing along with its validation by acetylation and homology-based search helped to identify the allergens from PS pollen proteome. Allergens causing cross-reactivity with coconut pollen have been unearthed and the allergenic potency of PS pollen extract has been described using in vitro allergenicity tests.

2. Materials and Methods

2.1 Pollen collection and preparation of antigenic extract:

Whole inflorescence containing clusters of flowers arranged on a stem was collected from date palm trees; anthers were isolated, dried, crushed and passed through meshes of decreasing pore sizes to obtain pure pollen. Defatting of pollen grains was done in diethyl ether. Crude pollen extract was prepared by incubating the defatted pollen grains in 0.1 M Phosphate Buffer (2:3 w/v) overnight at 4°C with constant stirring. The supernatant was used as an antigenic extract which was further diluted in 1:50(w/v), passed through the 0.22µm filter and sent to the clinic for Skin Prick Testing (SPT).

2.2 Selection of Patients:

Patients complaining of mild to severe seasonal allergies, bronchial asthma visiting Mediland Diagnostics, Kolkata, India was included for preliminary screening through SPT by pricking on the ventral side of the arm with a sterile lancet containing the PS pollen extract. Wheal diameter was monitored after 15 minutes and graded in +1 to +4 according to increasing allergenicity. Histamine diphosphate and phosphate buffer saline were used as a positive and negative control respectively. Smokers, patients suffering from immunodeficiency disorders and undergoing immunotherapy trials were excluded from the study. The entire study was approved by the ethical committee of the concerned institute. Ten patients who had >+3 SPT, high total IgE and substantial history of palm pollen allergy were recruited to provide blood for the experimental purpose after obtaining an informed consent. 5ml of blood was collected, serum separated and stored in -80°C until use. Sera from four non-atopic patients were taken as healthy controls (Supplementary Table 1).

2.3 1D, 2D Sodium Dodecyl Sulfate-Polyacrylamide Gel Electrophoresis (SDS-PAGE)

100mg defatted pure pollen was ground into fine powder in liquid nitrogen and mixed with 4ml chilled acetone containing 10% TCA(w/v), 1% DTT and 10 μ l/ml protease inhibitor cocktail (Sigma-Aldrich, USA). The slurry was vortexed mildly for 30 min at 25 °C and incubated in -20 °C overnight. After centrifuging for 20 minutes at 20000g the pellet was washed thrice with chilled acetone containing 1% DTT and air dried. The dried pellet was reconstituted in Isoelectric focussing buffer (7M Urea, 2M Thiourea, 2% CHAPS), vortexed mildly and incubated overnight at -20°C. Final protein was extracted by centrifuging at 20000g for 20 minutes at 4°C and estimated by Bradford reagent method.

For 1D PAGE 150 μ g of pollen protein in phosphate buffer was run in each well of 12% SDS-PAGE. For 2D PAGE 150 μ g of total pollen protein was reconstituted in 125 μ l of isoelectric focussing buffer containing 40mM DTT, 0.75% (v/v) IPG buffer (pH 4-7, GE life sciences) and 0.002%(w/v) bromophenol blue. Passive Rehydration loading was done in 7cm immobilized dry strip gels of pH 4-7 (GE Lifesciences) on a re-swelling tray overnight at room temperature. Isoelectric focussing was done in Ettan IPG phor 3 system (GE life sciences). Strips were thereafter equilibrated in equilibration buffer 1(6M urea, 75 mM Tris-HCl, pH 8.8, 29.3 % (v/v) glycerol, 2% SDS and 1% (w/v) DTT) and equilibration buffer 2 (same as equilibration buffer 1, replacing DTT with 2.5% (w/v) iodoacetamide) each for 15 minutes. The strips were finally laid on a 12% SDS-PAGE for 2D gel electrophoresis. Gels were stained in Coomassie Brilliant Blue R 250 and scanned in Gel Doc (Biorad, USA).

2.4 1D, 2D Immunoblot

Both 1D and 2D electrophoresed gel was electrotransferred onto PVDF membrane by semi-dry transfer at 1.2mA/cm² for 1 hour, blocked in Tris-buffered Saline+0.05% Tween 20 (TBST) containing 3% BSA for 2 hours at 4°C and then incubated in patient sera with primary antibody at 1:10 dilution in blocking solution overnight at 4°C. Secondary antibody

incubation was done in anti-human IgE alkaline phosphatase tagged produced in mouse (Sigma) at 1:1000 dilution for 3hrs at room temperature. IgE reactive bands/spots were visualized by Nitro-Blue Tetrazolium/5-bromo-4-chloro-3-indolyl phosphate (NBT/BCIP). Transfer of spots to the membrane was ensured by Ponceau S staining. Allergenic spots were identified by using Image master 2D platinum (GE life sciences) as well as visually by comparing 2D gel, immunoblot, and Ponceau S stain. 2D gel for blotting was done in parallel under identical experimental conditions to ensure that the protein spots exist in exactly the same coordinates of the gel. 2D gels and immunoblot were done in triplicates to minimize inter-gel variation. Another similar membrane was immunoblotted with pooled sera from four control patients.

2.5 Glycoprotein staining and deglycosylation

Glycoprotein staining of 2D gels was done using glycoprotein staining kit (G biosciences, USA) according to manufacturer's instruction. For deglycosylation, the 2D blotted membrane was incubated in 10 mM sodium acetate buffer (pH 5.0) containing 50 mM sodium metaperiodate for 3 hrs in the dark. Periodate was inactivated by sorbitol (0.015mol/L) with a drop of ethylene. The membrane was then incubated with sodium borohydride (1 mg/ml) for 12 hrs at 4°C and acidified with 0.001% acetic acid [8]. IgE reactive spots that do not contain glycan binding epitopes were detected with pooled patient sera as described before for immunoblotting.

2.6 In-gel digestion and MALDI-TOF/TOF

In-gel digestion was done following the protocol described previously [9]. In brief, desired IgE reactive protein spots were manually excised, destained in 1:1 ratio of 50mM sodium bicarbonate/ethanol followed by dehydration in acetonitrile. Gel pieces were then reduced in

10mM DTT and alkylated with 55mM iodoacetamide followed by rehydration and dehydration alternatively in 50mM sodium bicarbonate and acetonitrile. In-gel digestion was performed with 12.5ng/ μ l of trypsin (Promega, USA) overnight at 37°C. Peptides were extracted using 3% Trifluoroacetic acid (TFA) and 30% acetonitrile by vortexing vigorously for 20 min after removing trypsin. The extracted peptides were vacuum dried (Savant, Thermo Scientific, U.S.A.) and reconstituted in 5 μ l of 0.1% TFA and acetonitrile (1:1v/v). 2.5 μ l of the peptide was mixed with equal volume of 0.5 mg/ml of α cyano 4 hydroxycinnamic acid matrix (Bruker) and spotted on MTP ground steel plate (Bruker) in duplicates. Calibration was done using calibration standard II mixture (Bruker). Manual acquisition of spectra was done using Autoflex II MALDI-TOF/TOF in linear mode equipped with a pulsed nitrogen laser (λ = 337 nm, 50 Hz) at 55% power in the positive ion mode. Fragment ions were generated by hitting 4000 to 8000 laser shots on each of the dried spots. Peaks having a high signal-to-noise ratio were manually selected from each MS data and subjected to MS/ MS. The precursor and daughter ion spectra were analyzed using SNAP algorithm of Flex analysis version 3.4 after baseline subtraction and smoothing. Keratin, matrix ion, and trypsin autolysis peaks were removed using internal calibration of Flex analysis.

2.7 Database Search and analysis

Biotoools 3.2 (Bruker Daltonics) was used to transfer the selected MS/MS peaks to an in-house MASCOT search engine version 2.4.1. (<http://www.matrixscience.com>). Identification of the proteins was performed using the following specifications: Missed cleavage of one, precursor ion mass tolerance: 1.2 Da and fragment ion mass tolerance: 0.5 Da, Variable modification: methionine oxidation, fixed modification: carbamidomethylation of cysteine, taxonomy: Viridiplantae. Search was done in NCBI nr database against 106762850 sequences

and 39119668168 residues as on 31st October 2016. Results from MASCOT search were considered to be reliable if at least two peptides were identified from a single protein with scores above significance ($p < 0.05$) concluding significant homology or identity. The spots which could not be identified in the conventional database searching were *de novo* sequenced manually followed by a homology based search [10]. Sequence proposals generated through *de novo* sequencing from each peptide was used for homology-based search in MS-BLAST (<http://genetics.bwh.harvard.edu/msblast/>) [11] keeping the default settings: Program Blast 2p; non-redundant database (nr 95 2014_01); matrix as PAM30; “Expect” as 100. The result was determined in the positive hits according to the highest score in each search.

2.8 N-terminal acetylation

Dried tryptic peptides after extraction were diluted in 5 μ l of acetylation buffer (methanol containing 3% triethylamine and 12% acetic anhydride), incubated at 25 °C for 10 min and vacuum dried [12]. Peptides were again reconstituted in 5 μ l of 0.1% TFA and analyzed in MALDI-TOF/TOF according to the procedure described earlier.

2.9 Specific IgE Enzyme-linked Immunosorbent Assay (ELISA)

50 μ l of antigenic extract (100ng/ μ l) in 0.1 M Phosphate buffer saline (PBS, pH 7.2) was coated in wells of ELISA plate (Nunc, USA) overnight at 4°C and blocked with 1% BSA in PBS-Tween20. After washing, wells were incubated with 50 μ l of individual patient sera diluted 1:10 in blocking solution at 4°C overnight followed by incubation in monoclonal anti-Human IgE alkaline phosphatase tagged produced in mouse diluted 1:1000 in blocking solution at room temperature for 3 hours. IgE reactivity was detected colorimetrically by the pNPP substrate (Sigma). Absorbance values were recorded in ELISA reader (LabSystems Multiskan, Finland) at 405nm [13].

2.10 ELISA inhibition:

Serum pools of PS pollen positive patients were preincubated with 0.1, 1, 10 and 100, 1000 ng of crude PS pollen protein or coconut pollen extract or purified coconut major allergen as inhibitors. These sera were confronted with plate-bound PS pollen antigenic extract at 100 ng/μl and IgE titers were calculated similarly as done for ELISA. Percentage inhibition was calculated by

$$1 - \frac{(\text{OD}_{405} \text{ of sample with inhibitor})}{(\text{OD}_{405} \text{ of sample without inhibitor})} \times 100$$

2.11 Stripped Basophil Histamine release assay

Histamine release assay was done according to the method described previously [14]. In brief, Peripheral blood mononuclear cells (PBMC) was collected from nonallergic donors by Ficoll Paque (GE Lifesciences). Basophils were stripped off their bound IgEs by incubation in the lactic acid buffer for three minutes followed by washing in 4-(2-hydroxyethyl)-1-piperazine ethane sulfonic acid (HEPES) buffer (pH 7.5) and resensitized with 150 μl sera from each of the allergic patients at 37 °C for 90 min. The cells were stimulated with 1 μg of crude allergen extract in HEPES buffer for 1 h at 37 °C. The cell-free supernatant containing histamine was estimated by histamine assay kit (Immunotech, Beckman-Coulter). Cells sensitized with four non-allergic serum were taken as controls and BSA were used as a negative control. Histamine release was calculated as a percent of the total histamine content after subtracting spontaneous release.

2.12 Basophil activation Test:

PBMCs were extracted from the ten patients and suspended in stimulation buffer containing 1 μg/ml allergen [15]. Anti-human IgE was considered as a positive control and stimulation

buffer was used to determine basal level of CD63 in basophils. Cells were then incubated in 5µl of anti human CD63-FITC tagged antibody (Biolegend, CA, USA) for 15 minutes at 4°C. Unstained cells from the whole PBMC population were used for gating. The presence of CD63 was represented as a percentage of CD63 positive cells upon analyzing in FACS Calibur (Becton Dickinson, Franklin Lakes, NJ) using CellQuest™ Pro (BD) software at 488nm.

3. Results and Discussions

Current pharmacological treatments for allergic diseases can reduce the symptoms but are not able to cure the disease. The only therapy able to confer a long-lasting protection remains the allergen-specific immunotherapy (SIT). Identification of allergens is the primary step before making vaccines for SIT. The absence of proper information on allergenic components of PS pollen has inhibited diagnosis and development of efficient treatment regimes in highly prevalent PS pollen allergy in India. This study is the first attempt to identify the allergens from PS pollen that will pave the way for therapeutic interventions.

3.1 IgE reactivity to 1D and 2D SDS-PAGE profiled pollen proteins

Total protein from date palm pollen resolved in SDS-PAGE revealed more than 45 distinct bands in a region of 14 to 97 kD. Upon blotting with each patient-specific serum, nine IgE reactive bands were observed (Fig 1). A band at 20kD was immunoreactive to all the patients. A 33kD and 90kD band was IgE sensitive in 70% of the patients while others were moderately sensitive. At 50-55kD, a thick and diffused allergenic band was observed and more than one IgE reactive band was predicted in this region. This region was taken together and represented as a single band in 1D blot and all the patients were IgE reactive to this band. In two-dimensional gel profiling of PS pollen proteins, most of the protein spots were

concentrated in the 4-7 pI region in the 3-10 pI gel with only 139 ± 20 spots (Supplementary fig 1A) while 300 ± 25 spots were seen in the 4-7 pI gel (Fig 2A). So for further experimentation, gels were run in the 4-7 pI IPG strips where 12 IgE reactive spots were identified using pooled patient sera. Most of the allergens were concentrated in the 5-7 pI region and in the molecular weight range of 19-97 kD (Fig 2B). No allergenic spots were observed in control immunoblot (Supplementary Fig 1B).

3.2 Glycoprotein staining and deglycosylation

IgE binding epitopes are often found in the glycan part of glycoprotein allergens, triggering T_H2 responses, IgE production and also known as Cross-reactive Carbohydrate Determinants (CCD) which are responsible for cross-reactivity across a wide range of glycoprotein allergens [16]. To determine the presence of date palm pollen allergens that are glycoprotein in nature, protein resolved in the 2D gel was subjected to Periodic Acid–Schiff staining resulting in purple coloured spots. Four spots such as 1, 2, 10 and 11 were found to be glycoprotein in nature (Fig 2C). Allergens were then deglycosylated chemically by sodium metaperiodate, a strong oxidizing agent and probed with pooled patient serum to screen whether epitopes reside on the glycan or on the peptidic part. In spot 1, IgE reactivity was completely abrogated after deglycosylation suggesting that IgE epitopes are present in the glycan moiety (Fig 2D). In the other three glycoproteins, IgE binding epitopes were inferred to be peptidic in nature due to the retention of IgE activity after deglycosylation.

3.3 Identification of allergens by mass spectrometry

10 out of 12 allergens could be identified from date palm pollen based on homology with other plant species as shown in table 1. Conventional database search was first used as a screening step which helps identification of proteins that are highly homologous or identical

to them. This led to the identification of Beta-conglycinin, Carbonic anhydrase, S-Adenosyl methionine synthase and Beta-galactosidase with significant homology ($p < 0.05$). Peptides from these and other remaining spots which did not give significant results with database search were sequenced directly from MS/MS spectra manually and the generated sequence proposals were searched for similar or homologous proteins in MS BLAST. This resulted in the reliable identification of 7 more allergens. IgE reactive bands identified in the 1D gel were also subjected to mass spectrometry and could be correlated with the allergenic spots on the 2D gel as most of them were present in different molecular weights. Bands of 20, 21, 23, 33, 36, 44, 91 and 95kD correlated with spot number 2, 1, 3, 7, 4, 9, 10 and 11 respectively. Weak spectral quality due to low protein expression and non-significant homology hindered identification of spot 11 and 12. The frequency of IgE reactivity of the protein bands and their corresponding spots in the patients deduced from 1D immunoblot is represented in Table 2. At the 50-55kD region, four spots with very close molecular weights (5, 6, 8 and 12) were observed and are reflected in the 1D blot where no individual allergenic bands could be identified as proteins are less resolved in higher molecular weights. Though all patients were IgE reactive to this region which is contributed by the combination of all the four spots, their individual reactivity could not be calculated.

A single base change in the genomic sequence may lead to a change in an amino acid ultimately changing the mass of a peptide and causing the scores to steeply decrease in automated database searching. Confident identification of proteins from unsequenced organisms is possible when either it is searched in a database of an evolutionary close organism or falls under a highly conserved region where such SNPs are presumed to be negligible. Homology search relies on similarity which can be accomplished by allowing multiple mismatches in a sequence stretch [17].

In our study eight proteins showed homology with *Phoenix dactylifera*, a date palm species having same genus (Phoenix) and family (Arecaceae) probably because of both *P dactylifera*, *P sylvestris* are evolutionary close and genetically similar. Due to the agricultural importance of *P dactylifera* trees in Gulf Countries, its genome has recently been sequenced for fuelling biotechnological innovations in crop improvement [18]. For example, peptides from spot 2 (Fig 2A) showed significant homology with beta-conglycinin from *P dactylifera*. Interestingly three of the peptides were found to be present in the Cupin domain and since it is highly conserved, with negligible mutations, these could be identified easily through database search with a high score (Supplementary Fig 2). The Cupin domain containing proteins are a diverse superfamily of proteins containing a conserved beta-barrel domain. These proteins are mainly present as seed storage proteins known under different names such as 7S globulin, vicilin and are also reported as a major food allergen of nuts. They exist in homotrimers of 150-170kD and gets post-translationally processed to yield smaller allergenic subunits [19]. In this experiment, the protein spot was found to be present at a molecular weight of 20kD suggesting that it might undergo similar post-translational modifications and was also IgE reactive in all the patients (Table 2). This protein has also been identified as a major allergen from coconut pollen recently where a similar 20kD protein spot was allergenic [20]. Its function in pollen grains is yet unknown. Though it is a glycoprotein, its glycan part was not responsible for allergenicity similar to that observed in coconut pollen.

Spot 1(Fig 2A) was identified as pollen specific protein which has similarity with a major allergen from Olive (Ole e 1) pollen grains [21] and corresponded to the 21kD band in 1D blot showing moderate sensitivity. Ole e 1 belongs to a family of extracellular plant pollen glycoproteins which are about 145 residues long and contain six conserved cysteine residues but their biological function is yet unknown (Allfam: AF087; <http://www.meduniwien.ac.at/allfam/>). Similar allergens were also reported from family

Oleaceae, grass pollen and weed pollen, however, cross-reactivity in them does not extend beyond their families [22]. In Ole e 1, the carbohydrate moiety takes part in IgE binding [23] and a similar observation was also noticed in Spot 1 where allergenicity was completely removed after de-glycosylation.

Spot 4 (Fig 2A) showed significant homology with Alpha 1, 4 glucan synthase from *Elaeis guineensis*, another member of the Arecaceae family and was reactive in 50% of the patients (36kD band)(Table 2). This protein has been previously described as an allergen from *P dactylifera* pollen and responsible for cell wall synthesis which may assist in pollen tube growth during reproduction [24].

Isoflavone reductase (spot 7) was found to show intense immunoreactivity upon densitometric scanning compared to other allergens suggesting the presence of multiple IgE epitopes (Fig 2A, B). Isoflavone reductase belongs to a family of proteins that contains different NADP-dependent oxidoreductases involved in plant secondary metabolism (Allfam: AF123). Isoflavone reductase has been reported as an allergen in Olive [25], Coconut pollen [20] and Japanese cedar pollen [26].

S- adenosyl methionine (SAM) synthase was identified in spot 9 (Fig 2A) and was reported to be an allergen only in coconut pollen [20]. It is responsible for the formation of S adenosyl methionine which is required for polyamine biosynthesis. Spot 10 appeared as a string of three spots in the gel. All the three spots were found to have similarity with beta-galactosidase from *P dactylifera*. They were present near 90kD region and mass spectrometry analysis resulted in similar precursor and daughter ion spectra with an identical string of amino acids in each spot. This suggested that they could be isoforms of each other, all of which were allergenic in nature. Beta-galactosidase has been identified as an allergen only in palm pollen allergens such as *Phoenix dactylifera* [24] and Coconut [20] where similar

isoforms were observed to be IgE-reactive. It is also a glycoprotein but the IgE binding epitopes were peptidic in nature.

Four proteins were reported probably for the first time as novel allergens due to unavailability of similar allergens in allergen databases such as IUIS, Allergome, a Structural database of allergenic proteins (SDAP) or published literature. These are Carbonic anhydrase (spot 3), Xylose isomerase (spot 5), Glycerol synthase (spot 6) and ATP synthase alpha subunit (Spot 8) (Fig 2A). Carbonic anhydrase helps in the conversion of carbon dioxide and water to bicarbonate. Xylose isomerase catalyzes the interconversion of D-xylose and D-xylulose. Glycerol kinase is involved in triglycerides and phosphoglyceride synthesis and also participates in defense against the bacterium. Pollen-specific kinase genes are also potential signal transduction components of pollen–pistil interactions [27]. These four allergens are a new addition to the existing knowledge of IgE reactive proteins and need to be studied further to assess their individual contribution to allergenicity. Their allergenicity was predicted by an online software tool such as AlgPred (<http://www.imtech.res.in/raghava/algpred/submission.html>) which suggested all except ATP synthase alpha subunit be potential allergens using a Support Vector Machine module based on amino acid composition [28].

3.4 De novo sequencing and analysis.

Allergens are harbored in many pollen grains arising from species which are unsequenced. Methods such as whole genome sequencing or cDNA library screening for identifying allergenic proteins from these species are available but they are expensive and many laboratories do not have access to such technologies. *De novo* sequencing is another preferred method of protein identification when the database is absent and many allergens have been identified by this method such as in Coconut[20], Curvularia [13], fish [29], Hazel nut [30], *Lantana camara* pollen [31] to name a few. Automated *de novo* sequencing softwares are

often inefficient in identifying peptides with incomplete fragmentation and having low signal to noise ratio thus requiring manual intervention. In this study, we have used a manual method of *de novo* sequencing with spectra obtained from MALDI-TOF/TOF. As our dataset was small it was possible to interpret the spectra manually. *De novo* sequencing may be prone to errors as the incorrect assignment of 'b' and 'y' ions can generate a completely faulty sequence proposal. In many cases, chemical modifications have been used to validate the sequencing such as guanidination and sulfonation [32]. In this study, acetylation has been used to validate the *de novo* sequencing results. A study reported that *de novo* sequencing by acetylation gives better fragmentation patterns, simple to perform and recommended for sequencing small peptides from MALDI [33]. It has been used in the identification of outer membrane proteins from *Pseudomonas syringe* [34]. Acetylation adds an acetyl group to the amino terminal of each peptide thus shifting the 'b' ion series and the 'y_n' by 42 Da leaving the 'y' series intact if Lysine is absent in the sequence. *De novo* sequencing and its validation by acetylation of two peptides from spot 1 are shown in figure 3A, B. In the peptide 1016 Da, upon acetylation certain peaks were found to be increased by 42 Da which was inferred to be part of the 'b' ion series and helped to confirm a peptide stretch of N'EANPLGFLR C' showing similarity with pollen specific protein. Proline causes a low or no fragmentation towards the C terminal due to steric hindrance caused by the ring structure in ion transfer and often remains undetected leading to an insignificant score. In such cases, the residue mass of proline may be subtracted from the mass difference between two consecutive high-intensity peaks as shown in Fig 3A. Apart from the N-terminus, acetylation also occurs on the epsilon amino group of lysine. If there is a lysine termination, the y₁ will increase by 42 Da due to acetylation of Lysine and y_n will also increase by 42 Da due to acetylation at the N-terminus along with an increase of 42 Da in the b series. The total mass of the peptide increase by multiples of 42 Da according to the number of lysines present in the sequence in addition to

acetylation at the N-terminus. This effect is shown in acetylation of another peptide (1713Da) from spot 1 where an increase of total mass of peptide by 84 Da suggests additional acetylation at lysine at the C-terminus (Fig 3B). Acetylation is helpful in detecting peptides terminating with lysine since an y_1 fragment of 147 Da is of very low intensity and often remains undetected. Isobaric amino acids such as leucine and isoleucine cannot be differentiated by MALDI-TOF/TOF and are arbitrarily designated as leucine. *De novo* sequencing of all other spectra is provided in supplementary fig 5.

3.5 Specific IgE ELISA, Histamine release assay, and Basophil Activation Test

An elevated specific IgE response and histamine release confirms that the patient cohort under study is highly allergenic to PS pollen extract. Specific IgE ELISA showed 4 to 5.5 fold elevated IgE titers compared to healthy controls (Fig 4A).

Stripped basophil histamine release assay is a supplementary test to determine whether IgE is able to cross-link with each other and release histamine. Since histamine and other vasoactive amines are responsible for developing allergic phenotype, direct quantification of histamine release is a good measure of atopicity. It can be used as a confirmatory test to detect allergenicity of a protein or peptide. Basophils sensitized with PS pollen extract showed high histamine release in the range of 38-69% while control patients released less than 10% of histamine (Fig 4B).

Basophil activation test is another experiment to directly assay the atopicity of the proteins in recent years. After *in vitro* incubation with an allergen, the expression of CD63 on the basophil surface has been considered as the marker of basophil activation [35]. Basophils were quantified as percentage of CD 63 positive cells in the total population. Compared to healthy control, incubation of basophils from PBMCs of PS pollen allergic patients with crude extract showed a 3.3-4.5 fold increase in CD 63 expression (Fig 4C). Basal levels of

CD63 are non-significant compared to healthy controls implying biological basis of non-allergenicity. A representative FACS Plot of CD 63 upregulation upon allergen challenge in a patient is shown in Supplementary Fig 3. Significant positive correlation was observed with histamine between specific IgE and CD 63 positive cells (Supplementary Fig 4A, B). Some correlation was also observed with the number of allergenic bands in patients and specific IgE. Patients having a higher number of IgE reactive bands also had high IgE titers, Histamine release and CD63 markers suggesting the dominance of linear epitopes in PS pollen allergens.

Conformational epitope also contributes to IgE binding particularly in aeroallergens as they are presented to the immune system in the native state without denaturation as in the case of food allergens. In this study, the allergens are identified solely based on linear epitopes as they are analyzed under denaturing conditions. Individual allergens need to be purified, sequenced and further assayed to determine the presence of conformational epitopes which will provide insights into the contribution of each allergen to atopicity.

3.6 Identification of Cross-reactive allergens with coconut pollen

Strong cross-reactivity is observed mostly in closely related plants having a common progenitor. Some allergens are highly conserved such as pan-allergens and can cause cross-reactivity across vastly different species. Cross reactivity is supposed to be mediated by the presence of similar allergens between two species which shares common IgE binding epitope where they may have high sequence similarity. More than 60% of the patients immunoreactive to PS pollen visiting the clinic were also allergic to coconut pollen. Their close phylogenetic relatedness might have resulted in the generation of such cross-reactive allergens. Amongst the allergenic palm pollen grains, only IgE reactive proteins from coconut are known. In the quest of finding the cross-reactive allergens between coconut and date

palm, all the ten patients recruited in this study were essentially allergic to both the species as determined clinically by skin prick test.

Five allergens were found to be present commonly between PS and coconut pollen grains and are presumably responsible for cross-reactivity. These include Vicilin like protein, Isoflavone reductase, Alpha 1, 4 glucan synthase, SAM synthase and beta-galactosidase which account for more than 40% of the allergens from PS pollen. Interestingly most of the peptides that were identified from these proteins in PS pollen were identical to those obtained from coconut pollen through mass spectrometry [20]. This proves that both coconut and date palm are genetically very similar and cross-reactivity between them may be caused by similar IgE binding epitopes. Vicilin like protein was found to be allergenic in all patients suggesting it to be the major cross-reactive allergen. Isoflavone reductase and beta-galactosidase identified as 33 and 90kD band respectively in 1D blot were found to be allergenic in 70% of the PS allergic subjects while the other two were allergic to at least 50% of the patients (Table 2). The five cross-reactive allergens could be used in diagnostics to differentiate patient cohorts with both coconut and date palm sensitivity from true date palm pollen allergic subjects. The other allergens identified can act as markers for true date palm pollen allergy. The compartmentalization of the determined allergens in this study will have tremendous implications in immunotherapy as the cross-reactive group of patients will have a different treatment regime compared to true date palm pollen allergic patients. The cross-reactive allergens will have the potential to develop into better drug candidates targeting a wide range of patients.

To ascertain the extent of cross-reactivity of PS pollen extract, IgE inhibition was done with self-antigen, coconut pollen extract and purified native vicilin-like protein, the major allergen from coconut. A dose dependent inhibition of IgE binding to PS pollen extract was noticed when pooled sera were incubated with increasing concentration of these three antigens.

Approximately 9 ng of crude PS protein extract (Self-inhibition), 21ng of Coconut pollen extract and 230 ng of purified allergen was responsible for 50% IgE inhibition (Fig 4D). It could be suggested that vicilin-like protein may hold the promise for immunotherapy in patients cohorts sensitive to PS and coconut pollen allergy. Low concentration of 50% IgE inhibition values of coconut pollen extract suggests that the five allergens contribute significantly to cross-reactivity.

4. Conclusion

In this study, ten novel allergens from *Phoenix sylvestris* pollen were identified for the first time using *de novo* sequencing and homology-based proteomics. The *de novo* sequenced peptides can be used in designing degenerate primers for retrieving the coding sequence of the allergens thus paving the way for a generation of recombinant allergens. This can assist in understanding the three-dimensional structure of the allergens, mapping their epitopes, dissecting the molecular basis of allergenicity and thus the development of hypoallergenic vaccines. Allergens responsible for cross-reactivity with coconut pollen have been elucidated. This will enable clinicians for minimizing the number of allergens to be used in immunotherapy for such cross-reactive patients and at the same time in obtaining more precise and efficacious dosing for better results.

Acknowledgements

The study is supported by research funding from University Grants Commission (Grant Number: F.2-8/2002(SA-I)) and Department of Science and Technology, Govt. of India. Mr. Bidhan Singh from CRF, IIT Kharagpur and Amarendra Nath Biswas from Central Instrumentation Facility at Bose Institute are acknowledged for technical support in mass spectrometry. Authors are thankful to Jadab Ghosh, Kaberi Ghosh, Chanchal Chakraborty and Soumyo Gupta for their assistance in daily experiments.

Conflict of interest statement

Authors would like to declare that there are no conflicts of interest.

Author contributions:

BS performed the experiments. NP provided blood samples from allergic patients. SGB conceived the experiment and supervised the work.

ACCEPTED MANUSCRIPT

References

1. Blomme K, Tomassen P, Lapeere H, Huvenne W, Bonny M, Acke F, Bachert C, Gevaert P. Prevalence of allergic sensitization versus allergic rhinitis symptoms in an unselected population. *Int Arch Allergy Immunol*, 2012; 160(2):200-207.
2. Chakraborty P, Chowdhury I, Gupta-Bhattacharya S, Chanda S, Gupta S, Sengupta, DN. Clinicoimmunologic studies on *Phoenix sylvestris* Roxb. pollen: an aeroallergen from Calcutta, India. *Allergy*, 1999; 54(9): 985-989.
3. Chowdhury I, Chakraborty P, Gupta-Bhattacharya S, Chanda S. Allergenic relationship among four common and dominant airborne palm pollen grains from Eastern India. *Clin Exp Allergy* 1998; no. 8: 977-983.
4. Huertas, A. J., López-Sáez, M. P., & Carnés, J. (2011). Clinical profile of a Mediterranean population sensitized to date palm pollen (*Phoenix dactylifera*). A retrospective study. *Allergologia et immunopathologia*, 39(3), 145-149.
5. Kwaasi, A. A. A., Parhar, R. S., Tipirneni, P., Harfi, H., & Al-Sedairy, S. T. (1993). Major allergens of date palm (*Phoenix dactylifera* L.) pollen. *Allergy*, 48(7), 511-518.
6. Asturias, J. A., Ibarrola, I., Fernandez, J., Arilla, M. C., Gonzalez-Rioja, R., & Martinez, A. (2005). Phod 2, a major allergen from date palm pollen, is a profilin: cloning, sequencing, and immunoglobulin E cross-reactivity with other profilins. *Clinical & Experimental Allergy*, 35(3), 374-381.
7. Ma, B., Johnson, R. De novo sequencing and homology searching. *Mol Cell Proteomics*, 2012; 11(2), O111-014902.
8. Ghosh N, Sircar G, Saha B, Pandey N, Bhattacharya SG. Search for allergens from the pollen proteome of sunflower (*Helianthus annuus* L.): a major sensitizer for respiratory allergy patients. *PloS one*, 2015; 10(9), p.e0138992.

9. Shevchenko A, Tomas H, Havli J, Olsen JV, Mann M. In-gel digestion for mass spectrometric characterization of proteins and proteomes. *Nat Prot*, 2006; 1(6):2856-2860.
10. Kaufmann R, Spengler B, Lützenkirchen F. Mass spectrometric sequencing of linear peptides by product-ion analysis in a reflectron time-of-flight mass spectrometer using matrix-assisted laser desorption ionization. *Rapid Commun Mass Spectrom*, 1993; 7(10):902-910.
11. Shevchenko A, Sunyaev S, Loboda A, Shevchenko A, Bork P, Ens W, Standing KG. Charting the proteomes of organisms with unsequenced genomes by MALDI-quadrupole time-of-flight mass spectrometry and BLAST homology searching. *Anal Chem*, 2001;73(9):1917-1926.
12. Kulkarni HM, Ramesh V, Srinivas R, Jagannadham MV. Acetylating tryptic peptides enhances b ion intensity in MALDI-TOF/TOF: implications in peptide sequencing and identification of proteins in an Antarctic bacterium *Pseudomonas syringae*. *Proteomics Insights*, 2010.3, p.1.
13. Dey D, Saha B, Sircar G, Ghosal K, Bhattacharya SG. Mass spectrometry-based identification of allergens from *Curvularia pallescens*, a prevalent aerospore in India. *Biochim Biophys Acta* 2016; 1864(7):869-879.
14. Sircar G, Saha B, Mandal RS, Pandey N, Saha S, Bhattacharya SG. Purification, cloning and immuno-biochemical characterization of a fungal aspartic protease allergen rho 1 from the airborne mold *Rhizopus oryzae*. 2015. *PloS one*, 10(12), p.e0144547.
15. An S, Chen L, Long C, Liu X, Xu X, Lu X, Rong M, Liu Z, Lai R. Dermatophagoides farinae allergens diversity identification by proteomics. *Mol Cell Proteomics*, 2013; 12(7):1818-1828.

16. Halim A, Carlsson MC, Madsen CB, Brand S, Møller SR, Olsen CE, Vakhrushev SY, Brimnes J, Wurtzen PA, Ipsen H, Petersen BL. Glycoproteomic analysis of seven major allergenic proteins reveals novel post-translational modifications. *Mol Cell Proteomics*, 2015; 14(1):191-204.
17. Magno, J., Spirin, V., Santana Balbuena, T., Thomas, H., Adzhubei, I., Sunyaev, S., and Shevchenko, A. (2008) Protein identification pipeline for the homology-driven proteomics." *Journal of proteomics* 71, 346-356.
18. Hazzouri KM, Flowers JM, Visser HJ, Khierallah HS, Rosas U, Pham GM, Meyer RS, Johansen CK, Fresquez ZA, Masmoudi K Haider N. Whole-genome re-sequencing of date palms yields insights into the diversification of a fruit tree crop. *Nat commun*, 2015; 6.
19. Dunwell JM, Purvis A, Khuri S. Cupins: the most functionally diverse protein superfamily? *Phytochemistry*, 2004;65(1):7-17.
20. Saha, B., Sircar, G., Pandey, N. and Gupta-Bhattacharya, S. Mining novel allergens from coconut pollen employing manual de novo sequencing and homology-driven proteomics. *J Proteome Res*. 2015; 14(11): 4823-4833.
21. Asturias JA, Arilla MC, Gómez-Bayón N, Martínez J, Martínez A, Palacios R. Cloning and expression of the pan allergen profilin and the major allergen (Ole e 1) from olive tree pollen. *J Allergy Clin Immunol*, 1997; 100(3):365-372.
22. Lombardero, M., T. Obispo, B. Calabozo, A. Lezaun, F. Polo, D. Barber. Cross-reactivity between olive and other species. The role of Ole e 1-related proteins. *Allergy* 57 2002; no. s71: 29-34.
23. Batanero, E., Crespo, J. F., Monsalve, R. I., Martín-Esteban, M., Villalba, M., Rodríguez, R.. IgE-binding and histamine-release capabilities of the main

- carbohydrate component isolated from the major allergen of olive tree pollen, Ole e 1. J Allergy Clinical Immunology, 1999; 103(1), 147-153.
24. Postigo I, Guisantes JA, Negro JM, Rodríguez-Pacheco R, David-García D, Martínez J. Identification of 2 new allergens of Phoenix dactylifera using an immunoproteomic approach. J Investig Allergol Clin Immunol, 2009; 19(6):504-507.
25. Castro L, Crespo JF, Rodriguez J, Rodríguez R, Villalba M. Immunoproteomic tools are used to identify masked allergens: Ole e 12, an allergenic isoflavone reductase from olive (Olea europaea) pollen. Biochim Biophys Acta, 2015; 1854(12):1871-1880.
26. Kawamoto, S., T. Fujimura, M. Nishida, T. Tanaka, T. Aki, M. Masubuchi, T. Hayashi, O. Suzuki, S. Shigeta, K. Ono. Molecular cloning and characterization of a new Japanese cedar pollen allergen homologous to plant isoflavone reductase family. Clin Exp Allergy, 2002; 32, no. 7: 1064-1070.
27. Nouri E, Reinhardt D. Flowers and mycorrhizal roots—closer than we think? Trends Plant Sci, 2015; 20(6):344-350.
28. Saha, S., Raghava, G.P.S. AlgPred: prediction of allergenic proteins and mapping of IgE epitopes. Nucleic Acids Res 34.suppl 2 (2006): W202-W209.
29. Carrera, M., Canas, B., Vázquez, J., Gallardo, J. M. Extensive de novo sequencing of new parvalbumin isoforms using a novel combination of bottom-up proteomics, accurate molecular mass measurement by FTICR– MS, and selected MS/MS Ion monitoring. J Proteome Res, 2010; 9(9), 4393-4406.
30. Nitride, C., Mamone, G., Picariello, G., Mills, C., Nocerino, R., Canani, R. B., Ferranti, P. Proteomic and immunological characterization of a new food allergen from hazelnut (Corylus avellana). J Proteomics, 2013; 86, 16-26.

31. Ghosal, K., Saha, B., Gupta-Bhattacharya, S. Clinical and immune-proteomic approach on *Lantana camara* pollen allergy- a major health hazard. Allergy asthma Clin Immunol, 2016;12(33).
32. Samyn, B., Sergeant, K., Carpentier, S., Debyser, G., Panis, B., Swennen, R., Van Beeumen, J. Functional proteome analysis of the banana plant (*Musa spp.*) using de novo sequence analysis of derivatized peptides. *J Proteome Res*, 2007; 6(1), 70-80.
33. Samgina, T. Y., Kovalev, S. V., Gorshkov, V. A., Artemenko, K. A., Poljakov, N. B., Lebedev, A. T. N-terminal tagging strategy for de novo sequencing of short peptides by ESI-MS/MS and MALDI-MS/MS. *J Am Soc Mass Spectrom*, 2010; 21(1), 104-111.
34. Jagannadham, M. V., Abou-Eladab, E. F., Kulkarni, H. M. Identification of outer membrane proteins from an Antarctic bacterium *Pseudomonas syringae* Lz4W. *Mol Cell Proteomics*, 2011. 10(6), M110-004549.
35. Sanz M, Gamboa P, De Weck AL. In vitro tests: basophil activation tests. In *Drug Hypersensitivity* 2007: 391-402.

Figure legends:

Fig 1: 1D Immunoblot. P1-P10: Immunoblot with 10 patient sera. PS: protein profile of PS pollen extract in 12% SDS-PAGE. * indicates the position of the allergenic bands later confirmed by mass spectrometry. The IgE reactive bands are numbered alongside. C1-C4 immunoblot with sera from healthy patients(Control).

Fig 2: Identification of allergenic spots. (A) 2D gel electrophoresis of date palm pollen protein in 4-7pI range with marked allergenic spots.(B) Immunoblot using pooled patient sera.(C) In-gel glycoprotein staining. (D) Deglycosylation followed by immunoblotting with pooled patient sera. IgE reactivity of spot 1 disappeared after de- glycosylation suggesting presence of glycan binding epitopes.

Fig 3 *De novo* sequencing of spot 1 identified as Pollen-specific protein (A) *De Novo* sequencing of the non-acetylated peptide of m/z 1016Da. N-terminal acetylation showing a shift in the 'b' ions and 'y_n' ion by 42 Da suggesting monoacetylation at N terminus. (B) *De novo* sequencing of the non-acetylated peptide of m/z 1713 Da. Peptide derivatization by acetylation showing an increase of total mass of peptide by 84 Da. The 'y₁' ion is shifted by 42 Da due to acetylation of lysine at the C-terminus thereby increasing 'y' series by 42Da. 'b' series are increased by 42 Da due to normal acetylation at N-terminus. Total peptide mass changes by additional 42Da due to acetylation of a lysine residue along with 42Da due to N terminal acetylation. Generated sequence is shown alongside.

Fig 4: In vitro allergenicity assays using PS pollen antigenic extract:(A) Specific IgE ELISA. C1-C4 represent non-atopic control patients. (B) Box and whisker plot of stripped basophil histamine release assay, BSA was used as negative control (C)Box and whisker plot of

Basophil Activation Test. Anti-Human IgE was used as a Positive control, Stimulation buffer was used to measuring basal levels of CD 63. (D) ELISA inhibition assay to determine cross-reactivity with Coconut pollen. ELISA carried out with solid phase coated with PS pollen extract. PS pollen positive patient sera was preincubated with graded concentration of PS pollen extract(self antigen), coconut pollen extract and major allergen of coconut. Statistical Analysis: One-way ANOVA employing Tukey's multiple comparison Test. * means $p < 0.05$.

Table 1: Identification of allergens by MALDI-TOF/TOF. The underlined sequences are obtained from MASCOT search. The corresponding scores($p < 0.05$), query coverage are underlined. Other peptides are identified by *de novo* sequencing and MS BLAST scores are given alongside. The function of the proteins is collected from UniProt Knowledgebase (<http://www.uniprot.org/>). #: identified as an allergen for the first time. * indicates Glycoprotein allergen.

Table 2: IgE reactivity of the allergens with each of the patient sera deduced from the 1D western blots described in Fig 1. The molecular weight of the IgE-reactive band and its corresponding spot number in the 2D gel has been given. + indicates IgE reactive band while – indicates non-IgE reactivity for that particular band in the 1D western blot. Total numbers of patients reactive to each particular allergen have been represented in the last column. The 20kD band (Spot 2) is the major allergen. 4 spots (5, 6, 8 and 12) was very much close in terms of molecular weight as observed in 2D gel and could neither be identified as four different allergenic bands, nor their frequency of IgE reactivity was determined. It has been represented as a 50-55kD region according to the molecular weights of the spots.* represents allergens commonly found in Coconut pollen.

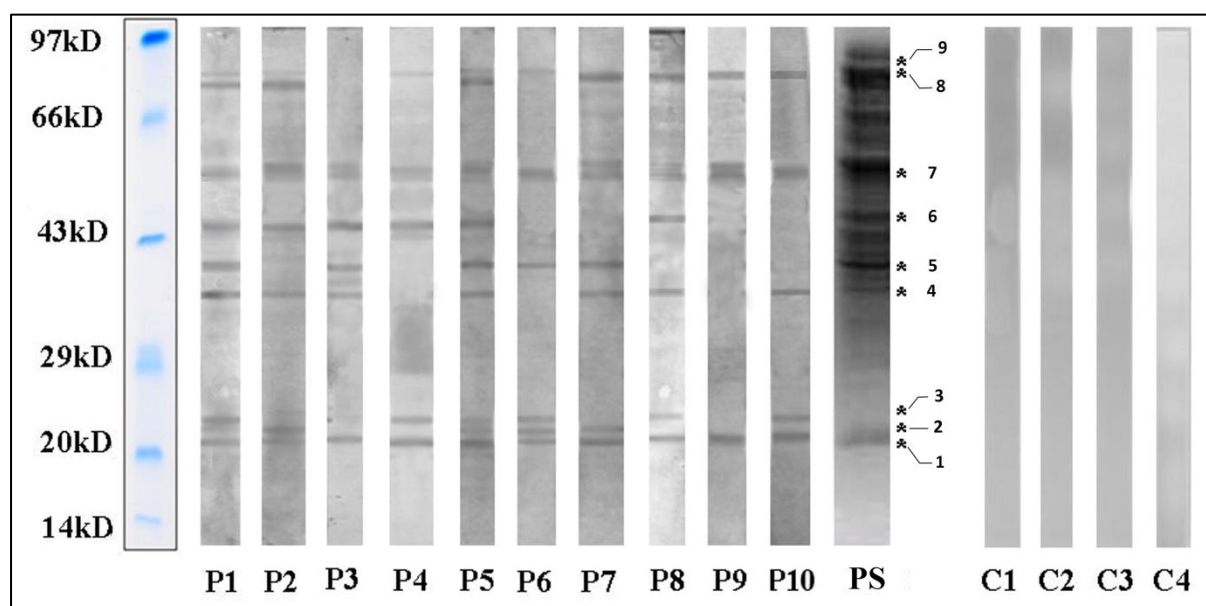
Fig 1

Fig 1: 1D Immunoblot. P1-P10: Immunoblot with 10 patient sera. PS: protein profile of PS pollen extract in 12% SDS-PAGE. * indicates the position of the allergenic bands later confirmed by mass spectrometry. The IgE reactive bands are numbered alongside. C1-C4 immunoblot with sera from healthy patients (Control).

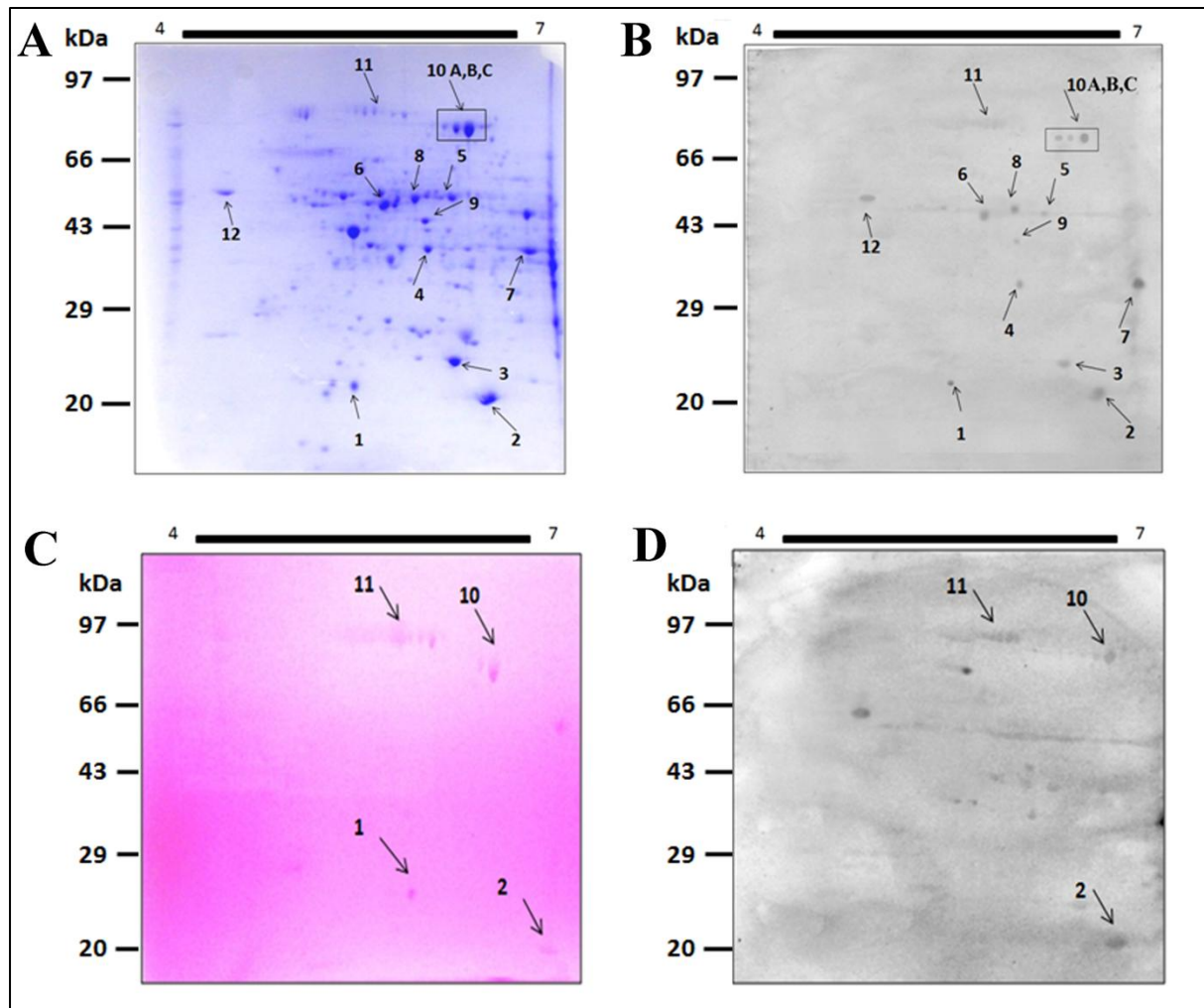
Fig 2

Fig 2: Identification of allergenic spots. (A) 2D gel electrophoresis of date palm pollen protein in 4-7pI range with marked allergenic spots.(B) Immunoblot using pooled patient sera.(C) In-gel glycoprotein staining. (D) Deglycosylation followed by immunoblotting with pooled patient sera. IgE reactivity of spot 1 disappeared after de- glycosylation suggesting presence of glycan binding epitopes.

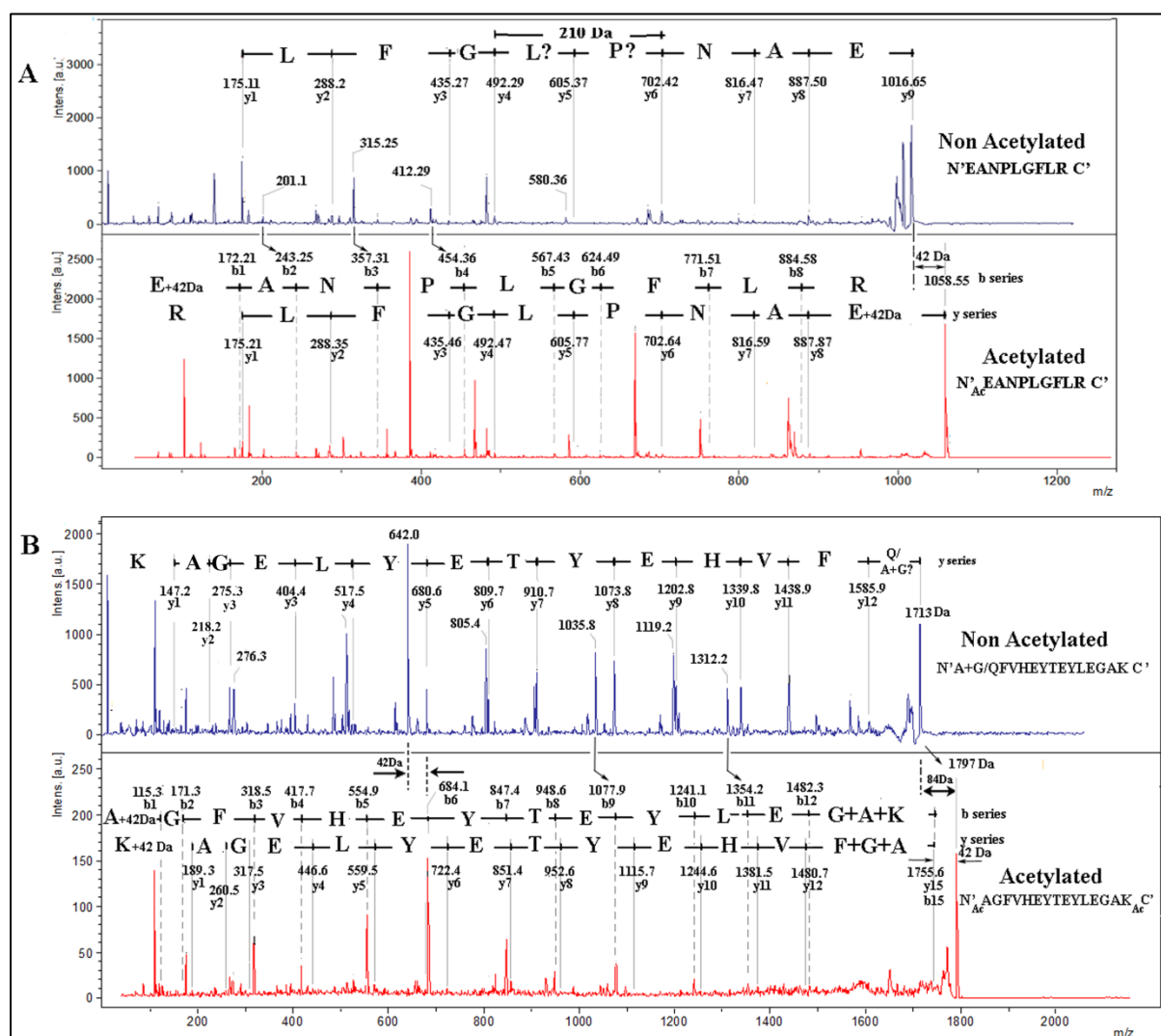
Fig 3

Fig 3 *De novo* sequencing of spot 1 identified as Pollen-specific protein (A) *De Novo* sequencing of the non-acetylated peptide of m/z 1016Da. N-terminal acetylation showing a shift in the 'b' ions and 'y_n' ion by 42 Da suggesting monoacetylation at N terminus. (B) *De novo* sequencing of the non-acetylated peptide of m/z 1713 Da. Peptide derivatization by acetylation showing an increase of total mass of peptide by 84 Da. The 'y₁' ion is shifted by 42 Da due to acetylation of lysine at the C-terminus thereby increasing 'y' series by 42Da. 'b' series are increased by 42 Da due to normal acetylation at N-terminus. Total peptide mass changes by additional 42Da due to acetylation of a lysine residue along with 42Da due to N terminal acetylation. Generated sequence is shown alongside.

Fig 4

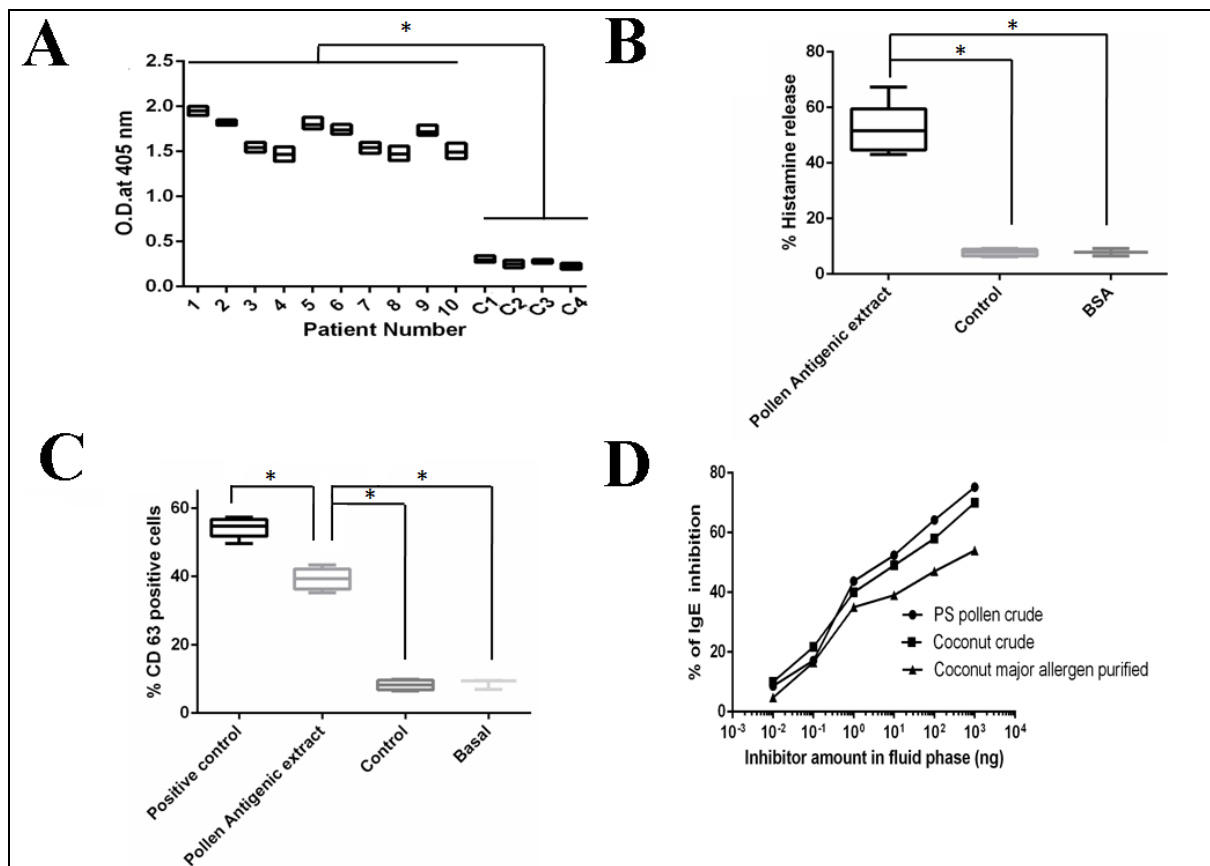


Fig 4: In vitro allergenicity assays using PS pollen antigenic extract:(A) Specific IgE ELISA. C1-C4 represent non-atopic control patients. (B) Box and whisker plot of stripped basophil histamine release assay, BSA was used as negative control (C)Box and whisker plot of Basophil Activation Test. Anti-Human IgE was used as a Positive control, stimulation buffer was used to measuring basal levels of CD 63. (D) ELISA inhibition assay to determine cross-reactivity with Coconut pollen. ELISA carried out with solid phase coated with PS pollen extract. PS pollen positive patient sera was preincubated with graded concentration of PS pollen extract(self antigen), coconut pollen extract and major allergen of coconut. Statistical Analysis: One-way ANOVA employing Tukey's multiple comparison Test. * means $p < 0.05$.

Table 1

Spot no.	Protein name	NCBI accession number	Organism	Theor. MW(kD) /pI	Exptl. MW(kD) /pI	Peptides	Score (p<0.05)	Method of Identification	Query coverage %	Unique peptide	Function
1	Pollen specific protein* C13 like	XP_008803750.1	<i>Phoenix dactylifera</i>	18.9/4.8	21/5.0	EANPLGFLR AGFVHEYTEYLEGAK	65 80	De novo	13.6	2	Pollen tube growth
2	Beta conglycinin* (Vicilin)	XP_008780644.1	<i>Phoenix dactylifera</i>	20/6.4	19/6.0	EAVILPTSPHR GPELAAAFGVSEEQLER NRPOFLVGASSLLHSMR YFPFCQIASNGAPLEFFGFITTSAR	53 141 78 83	MS/MS	36.8	5	Storage protein/ antimicrobial activity
3	Carbonic anhydrase #	XP_008782456.1	<i>Phoenix dactylifera</i>	18.9/5.3	23/5.27	TYPEFVEEALAK YAGVGSALIEYAVHHLK VVFDFQPGAEFTVR LKQYEHFK VFACADSR VLGHSR	54 62 107 35 60 41	MS/MS Denovo	54.1	7	Carbonate dehydratase activity
4	Alpha 1, 4 glucan synthase	AGE46030.1	<i>Elaeis guineensis</i>	39/6.2	36/5.88	MAGASVKPTPLKDELDIVIPTIR YDDMWAGWCTK YFGLMGDQGPIGR VPEGFDYELYNR GYPFSLR NEDVNLK LPYLWH	58 99 99 95 52 55 50	MS/MS De novo	22.2	7	Solute transport
5	Xylose isomerase#	XP_008796227.1	<i>Phoenix dactylifera</i>	46.6/5.7	50/5.0	FWHTFR LLNTDMKR WWCFHDR	55 49 70	De novo	6.0	3	Carbohydrate metabolism
6	Glycerol kinase#	XP_008779160.1	<i>Phoenix dactylifera</i>	56.6/5.15	54/6.0	LMENVDAVR GSLDKGTTSTR	68 63	De novo	4.4	2	Glycolysis
7	Isoflavone reductase	XP_010911620.1	<i>Elaeis guineensis</i>	33.7/6.0	32/6.01	NKGDLVDHS PSEFGNDVDR YTTVDEYLNK	94 70 46	De novo	10.1	3	Flavonoid biosynthesis
8	Atpase alpha subunit#	YP_005090378.1	<i>Phoenix dactylifera</i>	55.2/6.02	55/5.8	DVFYLHSR FQVDELGR ALGVPLDGR	62 59 59	De novo	3.1	2	Synthesis of ATP
9	s-Adenosyl methionine synthase	XP_008781205.1	<i>Phoenix dactylifera</i>	43.7/5.71	44/6.0	FVIGGPHGDAGLTGR TLFHLNPSGR YKNEGGAMVPLR	66 73 98	MS/MS De novo	8.9	3	Amino acid biosynthesis
10 a,b,c	Beta galactosidase*	XP_008811417.1	<i>Phoenix dactylifera</i>	93 /5.9	91/6.52	EIQNAGLYAILR SGSLHYPR EGNYDLLR WLLLFEEAKTRR KEHEVTLR	64 52 58 46 40	MS/MS De novo	4.7	5	Carbohydrate metabolism

Table 2

Band No./ MW	Corresponding spot in 2D gel	P1	P2	P3	P4	P5	P6	P7	P8	P9	P10	No. of patients Reactive
1/ 20kD	Spot 2*	+	+	+	+	+	+	+	+	+	+	10
2/ 21kD	Spot 1	-	+	-	-	+	+	+	-	-	-	4
3/ 23kD	Spot 3	+	-	-	+	+	+	-	+	-	+	6
4/ 33kD	Spot 7*	+	+	+	-	+	-	+	+	+	-	7
5/ 36kD	Spot 4*	+	-	+	-	+	+	+	-	-	-	5
6/ 44kD	Spot 9*	+	+	+	+	+	-	-	+	-	-	6
7/ 50-55kD	Spot 5,6,8,12	+	+	+	+	+	+	+	+	+	+	10
8/ 91kD	Spot 10*	+	+	-	-	+	-	+	+	+	+	7
9/ 95kD	Spot 11	+	-	-	+	+	+	-	-	-	-	4

Table 2: IgE reactivity of the allergens with each of the patient sera deduced from the 1D western blots described in Fig 1. The molecular weight of the IgE-reactive band and its corresponding spot number in the 2D gel has been given. + indicates IgE reactive band while – indicates non-IgE reactivity for that particular band in the 1D western blot. Total numbers of patients reactive to each particular allergen have been represented in the last column. The 20kD band (Spot 2) is the major allergen. 4 spots (5, 6, 8 and 12) was very much close in terms of molecular weight as observed in 2D gel and could neither be identified as four different allergenic bands, nor their frequency of IgE reactivity was determined. It has been represented as a 50-55kD region according to the molecular weights of the spots.* represents allergens commonly found in Coconut pollen.

Graphical abstract

