Supporting Information

Schulten et al. 10.1073/pnas.1300512110

SI Materials and Methods

Assembly of TG Pollen Transcriptome. Previously undescribed timothy grass (TG) pollen proteins that are potential T-cell antigens were identified by a combination of 2D gel/immunoblot analysis and cDNA sequence analysis of soluble pollen extract. Total RNA of TG pollen was isolated as described (Allergon AB) (1). In brief, 1 mg of ground pollen was homogenized in 20 mL of 4.2 M guanidine thiocyanate, 50 mM 2-bromoethanesulfonic acid, pH 7.2, and 4 mM EDTA followed by centrifugation at 15,000 × *g* for 15 min. Total RNA was prepared from the supernatant by using the TRIzol LS Reagent (Invitrogen) according to the manufacturer's instructions. The RNA was then analyzed by high-throughput sequencing on an Illumina Genome Analyzer. First, double-stranded cDNA was constructed from total RNA by using the Illumina TotalPrep RNA Amplification kit (Ambion). The cDNA was then converted into fragments of ∼200 bp by using NEBNext dsDNA Fragmentase (New England Biolabs). The ends were repaired, dA tails were added to the fragmented DNA by using NEBNext DNA Sample Preparation kit, and adaptors were added.

Sequencing was performed on an Illumina Genome Analyzer IIx (GAIIx). Briefly, adaptor-ligated cDNA was loaded into an Illumina flow cell. DNA was then bridge-amplified within the flow cell to generate millions of DNA clusters by using specific reagents and enzymes (Illumina Paired-End Cluster Generation Kit). The flow cell was loaded onto the GAIIx equipped with a paired-end module, and 72 sequencing cycles were performed to generate sequence in both directions by using Illumina Sequencing Kit v4. Replicate samples were run in seven of the eight lanes on the flow cell, producing >280 million raw sequence reads of 72 bp in length. Reads went through several preprocessing steps using the FastX toolkit (2) before they were assembled into contigs: (*i*) the 3′ terminal base was removed; (*ii*) low-complexity reads were removed; (*iii*) portions of reads downstream of a low-quality score were removed; and (*iv*) portions of reads corresponding to adapter sequencers were removed. The remaining reads were assembled into contigs by using Velvet (Version 1.0.15) (3). Because of the excessive memory requirements inherent to de novo sequence assembly, the reads for each lane were considered separately and were each run with five different values for the word size parameter (k = 21, 23, 25, 27, 29). We and others (4) have observed that different sets of contigs are obtained for each value for k. The contigs were further merged with Oases (Version 0.18.1; D. R. Zerbino, European Bioinformatics Institute, Hinxton, United Kingdom) into 1,764,158 putative transcripts. After redundancy reduction, 1,016,285 transcripts remained (including isoforms and other variants) with an average length of 245 bp and a maximum length of 6,884 bp.

Two-Dimensional Difference Gel Electrophoresis Analysis. The complete 2D difference gel electrophoresis (DIGE) analysis was performed by Applied Biomics. Briefly, TG extract was run on two 2D gels (3–10 pH range, 12% SDS–acrylamide gel), one was stained with Coomassie blue, and the other was blotted onto a nitrocellulose membrane. The membrane was then incubated with 5% dried milk in PBS/0.05% Tween to block nonspecific binding and subsequently probed with a serum pool from eight TG-allergic individuals at a dilution of 1:250. IgE and IgG binding were detected by using goat anti-human IgE and rabbit anti-human IgG (Sigma-Aldrich) and visualized by using Cy2-conjugated donkey anti-goat IgG and Cy5-conjugated donkey anti-mouse IgG antibodies (Biotium).

Pollen Protein Identification by Mass Spectrometry. Mass spectral protein identification was performed by Applied Biomics. Briefly, spots of interest were selected from the 2D blot, and the corresponding spots were identified on the stained SDS gel, cut out, and washed several times to remove staining dye and other inhibitory chemicals. The spots were then dried to absorb maximum digestion buffer. Dried 2D gel spots were rehydrated in digestion buffer containing trypsin. Proteins were digested in gel at 37 °C and then extracted from the gel with TFA extraction buffer. Subsequently, the peptides were desalted by using C-18 Zip-tips, (Millipore) mixed with an α-cyano-4-hydroxycinnamic acid matrix, and spotted into wells of a MALDI plate. Mass spectra of the peptides in each sample were obtained by using an Applied Biosystems Proteomics Analyzer. The spectra were compared with the amino acid sequences encoded by putative ORFs from the de novo-assembled TG pollen transcripts. All ORFs encoding for 15-amino acid or longer residues that had a >95% confidence hit as evaluated by the Mascot software package (Matrix Science) were considered hits (Dataset S1). The amino acid sequences encoded by these ORFs were clustered by using a custom script at a sequence similarity threshold of 90% to group together highly similar protein sequences. These sequences were encoded by different transcripts from our assembly, which could arise from splice variants, allelic variation between cells of TG pollen, and gene families with multiple members. Amino acid sequences that formed one cluster were assigned one putative protein ID (Dataset S1).

1. Wallner M, et al. (2009) Immunologic characterization of isoforms of Car b 1 and Que a 1, the major hornbeam and oak pollen allergens. *Allergy* 64(3):452–460.
2. Hannon GJ (2010) FASTX-Toolkit http://hannonlab.cshl.edu/fastx_toolkit/%3E.
3. Zerbino DR, Birney E (2008) Velvet: Algorithms for de novo short read assembly using de Bruijn graphs. *Genome Res* 18(5):821–829.
4. Surget-Groba Y, Montoya-Burgos JI (2010) Optimization of de novo transcriptome assembly from next-generation sequencing data. *Genome Res* 20(10):1432–1440.

Fig. S1. Production of IL-4, -5, -13, and all Th2 cytokines (IL-4, -5, and -13 combined) in response to TG extract, PHA, TG P20, and PUTGA P19. TG-stimulated cells were cultured in vitro with IL-2 for 14 d. Cells were restimulated with TG extract, the dominant known and previously undescribed TG peptide pools (TG P20 and PUTGA P19), and PHA. Th2 cytokine production was measured by ELISPOT to determine the best readout system.

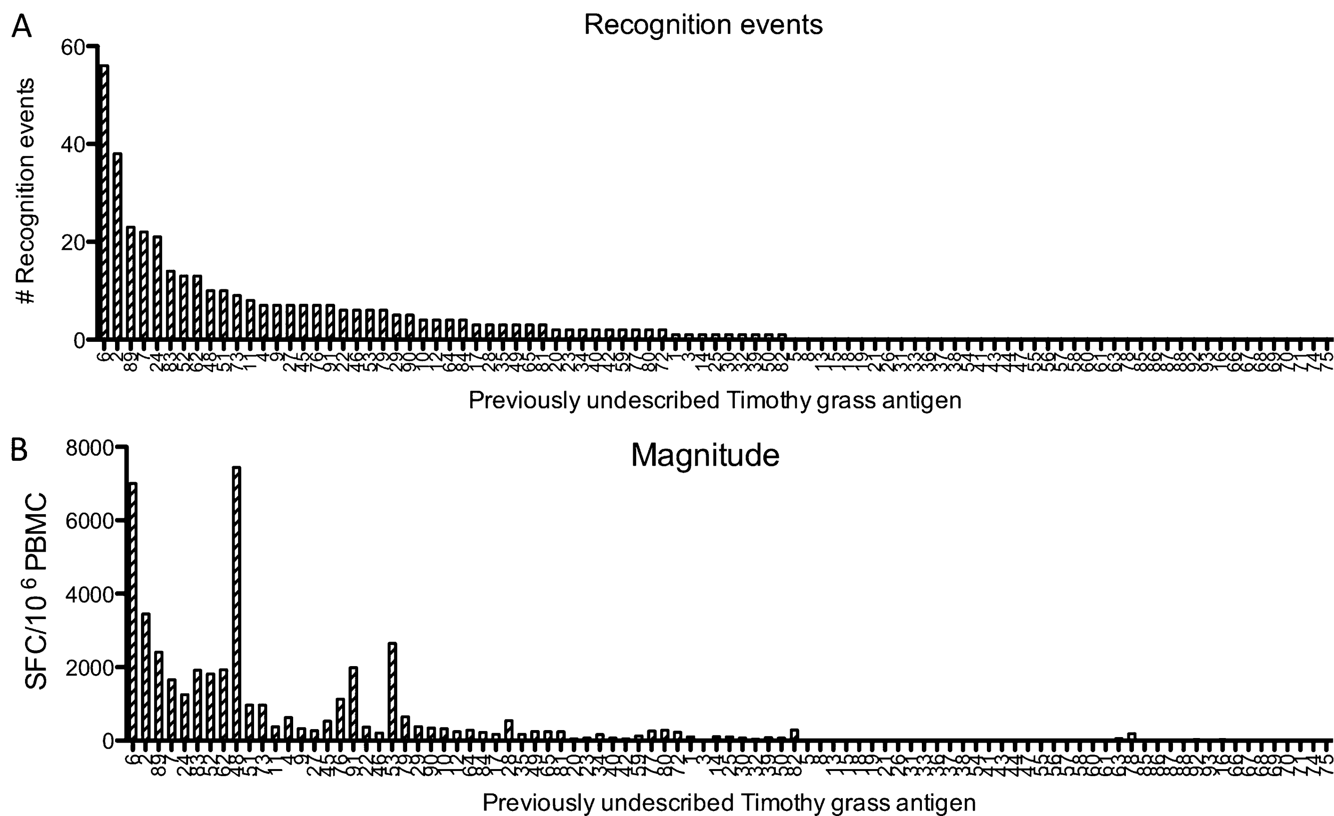

Fig. S2. Deconvolution of positive peptide pools to identify T-cell–reactive antigens and epitopes. IL-5 production from PBMC of allergic individuals in response to single peptides was measured. (*A*) Number of recognition events (defined as one peptide recognized by at least one patient with a magnitude of ≥20 SFC) per antigen tested. (*B*) Sum of the magnitudes of IL-5 responses against peptides from each positive antigen.

Table S1. Panel of 25 MHC II molecules for which peptide binding affinities were predicted

Locus	Allele
HLA DP	DPA1*01–DPB1*0401
	DPA1*0103–DPB1*0201
	DPA1*0201–DPB1*0101
	DPA1*0201–DPB1*0501
	DPA1*0301–DPB1*0402
HLA DQ	DQA1*0101–DQB1*0501
	DQA1*0102–DQB1*0602
	DQA1*0301–DQB1*0302
	DQA1*0401–DQB1*0402
	DQA1*0501–DQB1*0201
	DQA1*0501–DQB1*0301
HLA DR	DRB1*0101
	DRB1*0301
	DRB1*0401
	DRB1*0404
	DRB1*0405
	DRB1*0701
	DRB1*0802
	DRB1*0901
	DRB1*1101
	DRB1*1302
	DRB1*1501
	DRB3*0101
	DRB4*0101
	DRB5*0101

Table S2. Antibody reactivity and IL-5 production of PUTGAs and PUTGA-derived peptides

IgE	IgG	Antigens tested	Antigens positive	% positive	Peptides tested	Peptides positive	% positive	Total SFC	Peptides tested
-	+	24	14	58.3	173	47	27.2	14,127	173
+	-	13	9	69.2	118	36	30.5	3,124	118
+	+	16	12	75.0	187	93	49.7	14,157	187
-	-	30	15	50.0	256	74	28.9	11,607	256
n.d.	n.d.	10	2	20.0	88	11	12.5	1,197	88
Total	—	93	52	—	822	261	—	44,212	822

SFC, spot forming cell; n.d., not determined; — no total determined.

Other Supporting Information Files

Dataset S1 (XLSX)
Dataset S2 (XLSX)
Dataset S3 (XLSX)