Supplementary material to:

Identification of seasonal variations of antibodies against PR-10 specific epitopes can be improved using peptide-phage display.

Ramirez-Caballero L., Kny C., Treudler R., Simon J.C., Kern K., Jappe U., Szardenings M.

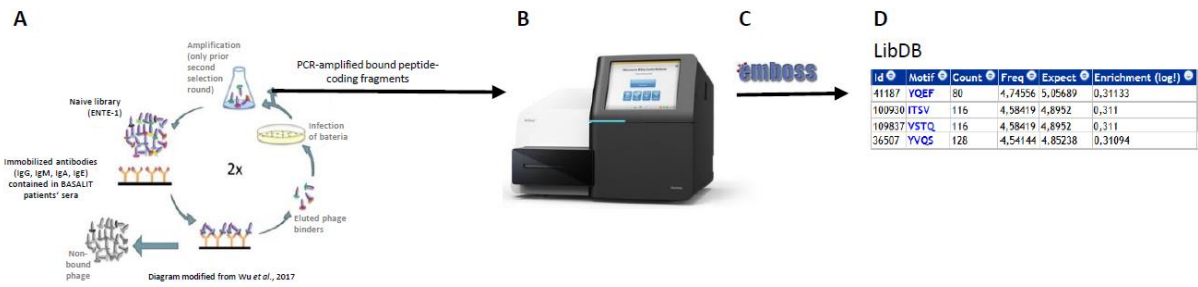**Generation of peptide phage display data**

Figure E1a. Peptide phage display data using patients' sera was generated as follows: **A. Phage display experiments:** Sera collected from birch pollen and soya-allergic placebo-treated patients (i.e. five samples $t_1 - t_5$ from each patient) were used for two consecutive rounds of selection experiments as described by Kern *et al.*, 2019. The starting library has a statistically even distribution with respect to the frequency of individual amino acids. **B. NGS sequencing:** PCR fragments of the peptide-coding genes are sequenced using illumina MiSeq®. **C. Data processing:** Sequences are further processed with tools of the EMBOSS software to remove sequencing errors. **D. Statistics of 4-mer epitope-related motifs in patient-datasets:** Sequencing data are uploaded into a database (LibDB), and further attention is paid to remove remaining detectable sequencing errors. One dataset of 4-mers for each selection round is generated. Their frequency is calculated taking into account the design of the naive library (expected value, [plog]) and the statistical occurrence in the dataset (frequency value, [plog]) resulting in an enrichment value (log). This is compared amongst the different datasets, also using the naïve library as reference. The analysis would not be viable based on trimers or pentamers due to a number of reasons. Motif analysis using trimers would yield in an excessive number of unspecific background sequences containing a given motif. Statistical analysis based on pentamers would need such a large amount of data, that the required number of binding clones is unrealistic. Only enriched epitope-related motifs are considered for further analysis by alignment of the sequences with these motifs (see Figure E1b).

ITSVE − Patient P2

t1 (sample collected at start of autumn 2012)	t4 (sample collected at end of spring 2013)

t1 (start of autumn 2012)

```
Cor a 1.01   KVAPQAITSVENVE-GNGGPGTIK
count=2      -----HITSVETYI-PYELVQ-Y-
count=2      -----VIFSVERIP-IKSPTK-I-
count=2      -----VSTSVEGLL-SPTRFS-I-
count=3      -----EETSVERLY-NPVPRV-C-
count=2      -----QITSVVALI-KPPDYV-F-
count=4      -----PITSDEWQV-CIDPNC-F-
count=2      -----TIPSVEPEL-AGEPVE-Y-
count=2      -----EIMSVEDAD-REHQYG-F-
count=2      -----YITSVHLLS-PCQAEA-C-
count=2      -GNMSSVTSVEEIC-LAI------
count=3      -----EILSVEKPY-LWVVGF-C-
count=2      -----YITSNEGIY-VPRSLI-I-
count=2      -----VIWSVEHEF-YCHKRP-D-
count=3      -----VKTSVERTC-RKVLLT-D-
count=2      -----HQTSVEDYC-VPGCYN-A-
count=2      -----IITSWEFNF-DQTLIE-C-
count=2      -----VITSVGFPF-YITQNS-F-
count=2      -----AITSVEW-----------
count=2      QSVTDAITEVELYI---------
count=3      -----PITSVEPIN-RWFTFP-F-
count=2      -----KITSVWNLN-DCPGDQ-Y-
count=2      -----SIHSVENWV-LQIYFS-D-
count=3      -----WIVSVENIF-DRVPPH-Y-
count=2      -DYGAGIHSVEN-----------
```

~ 34 out of 100,000 sequences retrieved from all permutations of ITSVE

Bet v 1-specific IgE concentration: 54.5 kUa/l
Cor a 1-specific IgE concentration: 40.3 kUa/l

t4 (end of spring 2013)

```
Cor a 1.01   KVAPQAITSVENVE-GNGGPGTIK
count=2      FCQQYAITWVELSI---------
count=6      -----AITSVIYPP-GESQFN-C-
count=4      -----AWTSVEYDP-GEVNTK-F-
count=5      -----QITSVGVWE-EWYQVH-C-
count=10     -----LITSVSKHD-FPIFCC-I-
count=3      -----TITSVNRWL-DCPCPV-F-
count=5      YWRWQHIYSVEC-----------
count=4      -----KITCVEHEE-FWIFGY-A-
count=2      QSKIRDITWVEFAA---------
count=8      -----EITSTENQH-LRLAES-Y-
count=2      -----FITSPEN---EPCYPDQSI
count=6      -----IITSGEANL-HWNQGK-F-
count=3      -----DGTSVENVV----KPPACN
count=11     -----LYTSVEDVI-RIQPCG-I-
count=4      TSYLDEPTSVEC-----------
count=5      -----KETSVEDTE-QRCCEQ-D-
count=4      -----AISSVEANY-ARE-AQEC-
count=2      -----APTSVELHN-KEVRPH-C-
count=4      -----HITSVLEVR-GK--RETCI
count=2      -----IITSVRPVN-GELVCY-C-
count=3      -----IIHSVENKL-HEIHDQ-I-
count=14     -----RIASVEVFA-KWIGRP-I-
count=4      -IHFSYIQSVEQDI-GIF------
count=4      --NNRSITSWEQQY-GCLI-----
count=2      -----HITSRELHE-PYYTPF-A-
count=4      ------ITSCEFLW-SLQRLH-CI
count=2      NVCDISITSWEYVQ-D--------
count=5      WVQVAVRTSVEN-----------
count=11     -----FITSHEHSQ-YITATP-N-
count=7      YYWSSWHTSVEFNH-DF-------
count=7      --DFNCRTSVESEK-YQQD-----
count=8      -----HRTSIWNLL-ICSVEF-D-
count=2      VPSWSAITSVYY-----------
count=14     -----FIWSVELEV-EPEKDN-D-
count=8      EEFCATIKSVEVDH-QF-------
count=21     -----FITSQEHPV-AEETLF-C-
count=3      -----GITSVHSFI-LPHIRN-N-
count=9      --WTHCITSHEPKE-VCGN-----
count=7      -----GITSNEYIA-FYHSLL-F-
count=3      RIRCKHIASVEYTQ-YY-------
count=3      -----NQTSVEQHQ-YWQQNI-D-
count=4      -----VIWSVEEQT-FCQPDC-A-
count=2      -----DIMSVESEH-HWQANI-C-
count=3      --QRLCKTSVEIWD-DPHA-----
count=5      -----FITSVTWIV-ARAPHS-Y-
count=7      -----IQTSVEEFQ-LGCLNY-C-
count=2      VSGCWPIQSVEIHL-FA-------
count=13     HHHLITIVEDAY-----------
count=12     HVAITPIGSVEY-----------
count=5      --IWWSITDVEQEA-EQLD-----
count=29     VFQSTSVECSREDH-FF-------
count=3      -----EITSAEYRN-LQDQEA-D-
count=4      YHRYVHCTSVEI-----------
count=10     YFLQIVITVVEI-----------
```

~ 106 out of 100,000 sequences retrieved from all permutations of ITSVE

Bet v 1-specific IgE concentration: 65.0 kUa/l
Cor a 1-specific IgE concentration: 43.7 kUa/l

Figure E1b. Typical examples for aligning enriched sequences containing motifs for the same potential epitope but from sera drawn at different seasons. Shown are only sequences containing four amino acid residues of ITSEV, i.e. *TSVE, I*SVE, IT*VE ITS*E ITSV* (where * represents any residue) and with in total five or more neighbouring residues identical to the allergen. Single sequences (i.e. sequences only occuring once) are not taken into account, because they may contain sequencing errors. Sequences were aligned against the allergen's primary structure and identical amino acid residues were highlighted. Typically, sequences are more frequently (count) observed in datasets obtained from serum with high IgE concentrations. This frequency was compared to sequences retrieved from the naïve library dataset, where only 6 out of 100,000 sequences were obtained using the same search. No sequence with more than two counts was observed in the naïve library dataset, showing a clear enrichment after selection rounds using patient sera.

H: Histidine
D: Aspartic acid
P: Proline
Y: Tyrosine
S: Serine
A: Alanine
E: Glutamic acid
V: Valine
I: Isoleucine
G: Glycine
T: Threonine
N: Asparagine

β-strand β2
Bet v 1: ISSVEN
Cor a 1: ITSVE

β-strand β6
Cor a 1: YTVIEG
Gly m 4: SVVG
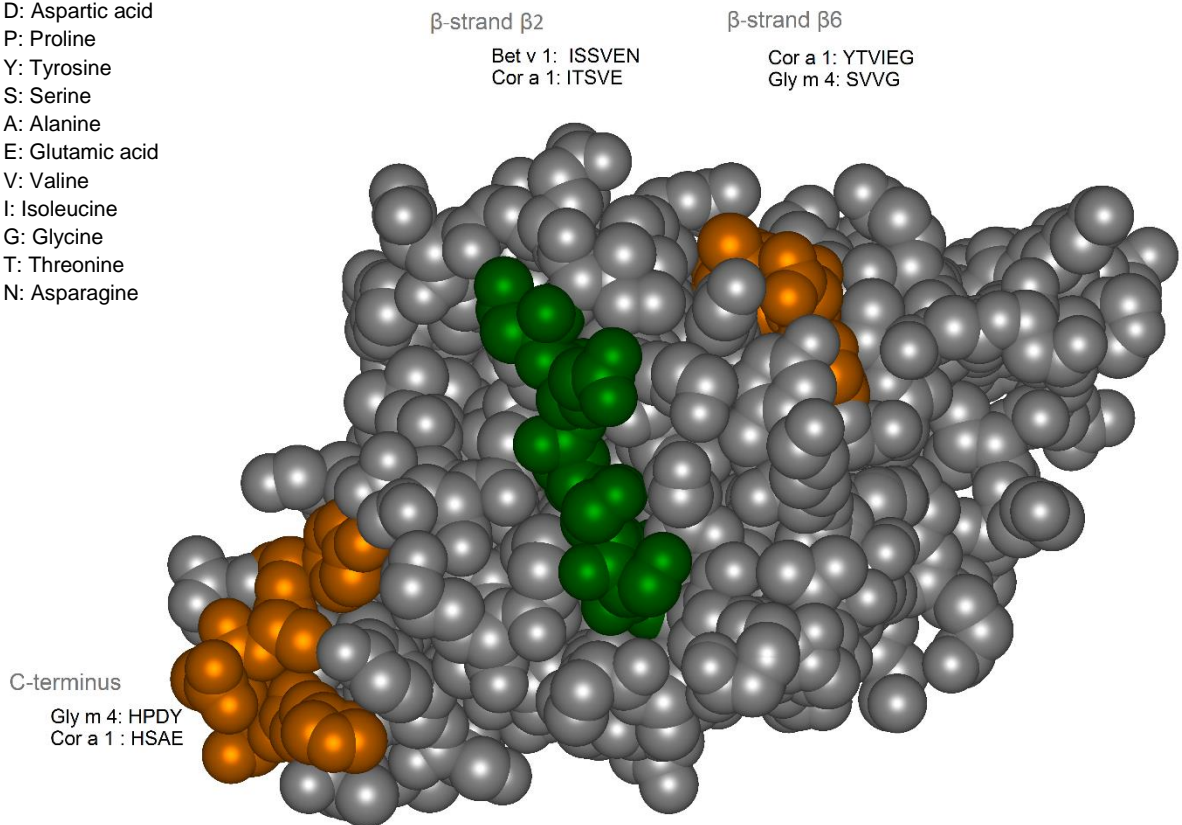
C-terminus
Gly m 4: HPDY
Cor a 1 : HSAE

Figure E2. Location of the PR-10 allergen-related motifs described in Figure 1 within the three-dimensional structure of the major birch pollen allergen Bet v 1 (PDB ID:1BV1). Motifs are located at β strand β2 (green), β strand β6 (orange, right) and C-terminus (orange, left). Similar motifs from Gly m 4 and Cor a 1 are indicated.
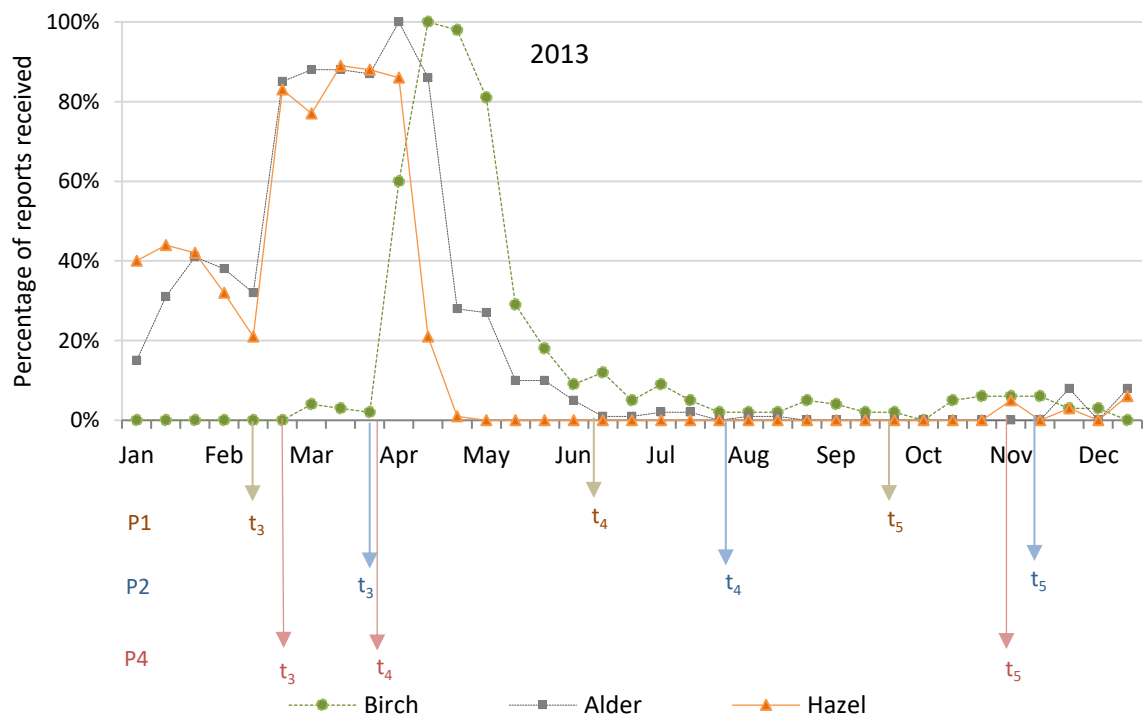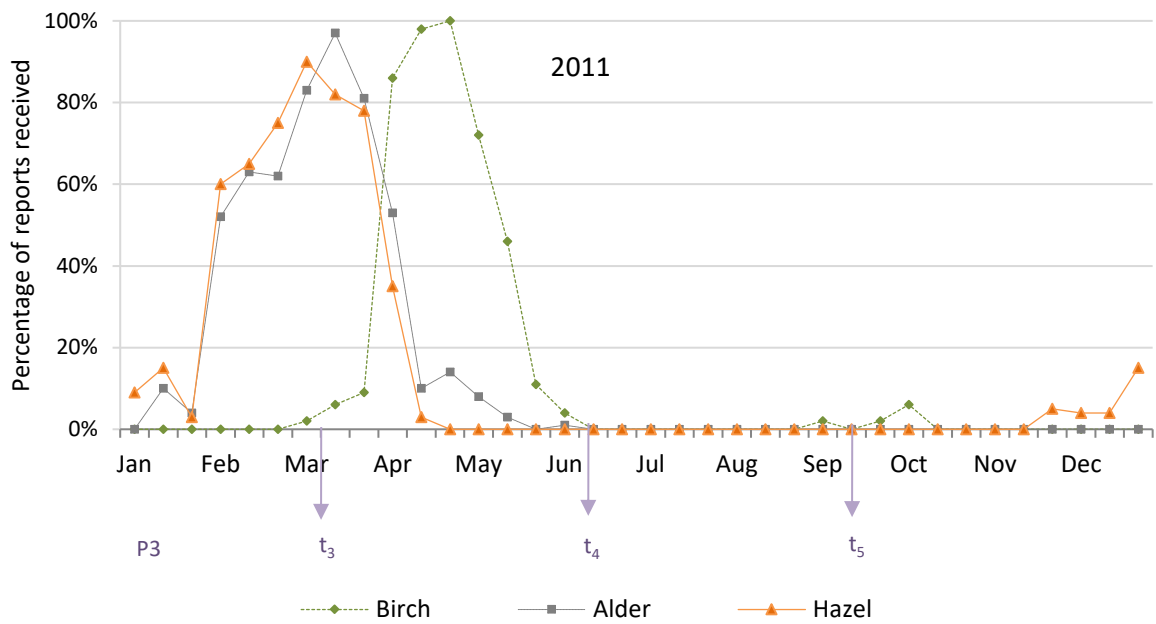
Figure E3. Received reports of birch, alder and hazel pollen exposure in Germany in 2011 (above) and 2013 (below). The gray arrows indicate when serum was collected, labeled by Patient ID and the corresponding time-point ($t_3 – t_5$). The daily average of pollen particles per square meter is calculated based on data from all measurement stations. The average of the measurements taken every ten days is represented along the X-axis. The percentage indicates the received reports in Berlin, Saxony, Saxony Anhalt, and Thuringia, i.e. the region where the patients were studied. Plots were created from records provided by the German Weather Service (*Deutscher Wetterdienst: https://www.dwd.de/DE/leistungen/pollen/pollenstatistik.html*).

Table E1. List of motifs with the largest variations of their enrichment values in serum collected over one year. 4-mer motifs identified in patients are indicated, according to their patient ID (P1 – P4) and their amino acid sequence. Motifs that are part of epitopes reported in the Immune Epitope Database are indicated by their identification number. When no epitopes containing the motifs are reported in IEDB, epitope candidates identified in-house from an ongoing larger study (unpublished data) are indicated instead. In this case, the amino acid residues and their position within the allergen are indicated.

Allergen (GenBank ID)	Motifs with largest variations in their enrichment values throughout all time-points [Patient ID - 4-mer motifs]	Corresponding epitope (IEDB ID)
Bet v 1 (Bet v 1.0101 CAA33887)	P1 - APQA, PQAI, QAIS; SVEN, VENI, ENIE P4 - VAPQ, APQA	35940, 141047, 191291, 191686
	P4 - VDEV, EVDH	21651, 181135
	P2 - RAVE, AVES	41567, 58813
Gly m 4 (CAA42646)	P2 - VFTF, FTFE, TFED, FEDE, EDEI	Not reported in IEDB (Epitope candidate $_2$GVFTFEDEI$_{10}$)
	P2 - KALV, ALVT	Not reported in IEDB (Epitope candidate $_{21}$KALVTDA$_{27}$)
	P2 - SVEN, VENV, ENVE P1 - VENV, ENVE, NVEG	913573
	P3 - ETKF, TKFV, KFVL P4 - LEDG, EDGE, DGET	580931
	P3 - SVVG	580708, 914272
	P3 - DAEP, AEPN, EPNQ, PNQD, NQDE	913251, 913938
	P3 - LLAH, LAHP, AHPD, HPDY	Not reported in IEDB (Epitope candidate $_{151}$LLAHPDYN$_{158}$)
Cor a 1 (Cor a 1.0101 CAA50327)	P2 - SVIP, VIPA, IPAA	Not reported in IEDB (Epitope candidate $_3$GVFNYEVETPSVIPAA$_{18}$)
	P1 - APQA, PQAI P4 - VAPQ, APQA	76332, 69411
	P1 - SVEN, VENV, ENVE	Not reported in IEDB (Epitope candidate $_{39}$ITSVENVEG$_{47}$)
	P4 - VDEV, DEVD	Not reported in IEDB (Epitope candidate $_{70}$ERVDEVDN$_{77}$)
	P3 - SVIE, VIEG, IEGD, EGDV, GDVL	Not reported in IEDB (Epitope candidate $_{84}$YTVIEGDVLG$_{93}$)
	P1 - GDHE, DHEI	Not reported in IEDB (Epitope candidate $_{124}$KGDHEINAEE$_{133}$)