

E. Kosman^a

A. Eshel^b

Y. Waisel^b

^a School of Mathematical Sciences and

^b Department of Plant Sciences,

Tel Aviv University, Tel Aviv, Israel

The 'Travelling Salesman Problem': A New Approach for Identification of Differences among Pollen Allergens

27414

Key Words

Allergens

Cross-reactivity

Clustering

Pollen

Skin tests

Travelling salesman problem

Abstract

Background: It is not easy to identify the specific plant species that causes an allergic response in a certain patient at a certain time. This is further complicated by the fact that closely related plant species cause similar allergic responses. A novel mathematical technique is used for analysis of skin responses of a large number of patients to several groups of allergens for improvement of the understanding of their similarity or dissimilarity and their status regarding cross-reactivity. **Methods:** The responses of 153 atopic patients to 42 different pollen extracts were tested by skin prick tests. Among the responses of patients to various extracts, a measure of dissimilarity was introduced and calculated for all pairs of allergens. A matrix-structuring technique, based on a solution of the 'Travelling Salesman Problem', was used for clustering of the investigated allergens into groups according to patients' responses. The discrimination among clusters was confirmed by statistical analysis. **Results:** Sub groups can be discerned even among allergens of closely related plants, i.e. allergens that are usually regarded as fully cross-reactive. A few such cases are demonstrated for various cultivars of olives and pecans and for various sources of date palms, turf grasses, three wild chenopods and an amaranth. **Conclusion:** The usefulness of the proposed approach for the understanding of similarity and dissimilarity among various pollen allergens is demonstrated.

Introduction

Pollen, spores and various other airborne particulate organic matter may invoke allergic responses in atopic people [1, 2]. The effects of such allergens are direct and, in most cases, also specific. However, because pollen and spores are carried by winds over long distances and because they appear in the air in a large assortment, it is not easy to identify the specific plant species that causes an allergic response in a certain patient at a certain time. This is further complicated by the fact that closely related plant species cause similar

allergic responses [3–6], and by the difficulties to distinguish between responses to specific types of pollen when a wide spectrum of airborne pollen is present [7–10].

The purpose of the following study was to broaden the application of a mathematical tool for the analysis of large sets of data of skin tests [10]. This method depicts the similarity, and dissimilarity, between examined allergens as revealed by the skin responses of a large number of patients. Such information is important for the improvement of the understanding of the similarity of various plant allergens and affirms their status regarding cross-reactivity.

KARGER

E-Mail karger@karger.ch
Fax +41 61 306 12 34
<http://www.karger.ch>

© 1997 S. Karger AG, Basel
1018–2438/97/1124–0371 \$12.00/0

This article is also accessible online at:
<http://BioMedNet.com/karger>

Correspondence to: Prof. Yoav Waisel
Department of Plant Sciences, Tel Aviv University
IL-69978 Tel Aviv (Israel)
Tel. 972 3 640 9189, Fax 972 3 640 9380
E-Mail waisel@post.tau.ac.il

Materials and Methods

A survey of population responses to various pollen allergens was carried out in 10 rural settlements in Israel. Skin prick tests were performed with autochthonous pollen extracts. The extracts were prepared either from pollen batches that were collected from individual trees (Date palms – *Phoenix dactylifera*) or from cloned plants of well-defined cultivars (pecans – *Carya illinoensis*; olives – *Olea europaea*; turf grasses etc.). Pollen of other plants was collected in Tel Aviv from the local populations of weeds. The autochthonous extracts were prepared as follows: clean pollen grains were sonicated in a sterile saline-glycerol buffer solution (pH 8.2; 2.8% NaHCO₃, 0.5% NaCl, 50% glycerol). The solution was centrifuged for 30 min at 7,600 g, and the supernatant was then filtered through a membrane with 45-μm pores. Protein content of the obtained extract was determined by the Lowry procedure, with BSA as a standard. After the protein content was determined, each of the autochthonous extracts was diluted to a standard concentration of 5 mg protein/ml solution. The extracts were stored at –20°C [for the detailed procedure of extract preparation, see ref. 11]. Though the tested allergens are typical for many of the Mediterranean countries, the method of analysis is suitable for any set of allergens. Concomitant tests were made with commercially available pollen extracts (produced by Dome/Hollister Stier, Spokane, Wash., USA and by Center Laboratories, Port Washington, N.Y., USA). Altogether, 42 allergens of different sources were tested on most of the 153 patients. Each of the patients was tested for at least 30 allergens. Only 2 allergens were not tested concomitantly on any of the patients (marked by '?' in fig. 1). The list of all allergens tested is presented in figure 1.

Skin prick tests were made on both arms of the patients, and the diameter of the developing wheals was measured and classified into 5 grades: 0: wheal diameter smaller than 2 mm; 1: wheal diameter 2–4 mm; 2: wheal diameter 4–10 mm; 3: wheal diameter 10–15 mm, and 4: wheals larger than 15 mm, mostly with developed pseudopodia. In order to identify patients with exceptional responses, histamine and saline-glycerol pricks were performed on each arm.

Results of the skin tests of all patients, based on the frequency of responses and their intensity (size of the wheal), were used for the calculation of the degree of dissimilarity between each pair of allergens, i.e. the so-called 'distances' between them. The distances were calculated on the basis of the intensity of responses of patient *t* to allergen *r* (*i_{rt}* = 0, 1, 2, 3 or 4, depending on the diameter of the wheal). A very sensitive patient who responds to many of the allergens tested is characterized by many positive *i_{rt}* values. The *i_{rt}* values equal zero in a nonatopic patient.

When a patient *t* is tested for 2 allergens, *r* and *s*, the measure *d_{rst}* = |*i_{rt}* – *i_{st}*|, which corresponds to the difference in diameter of the two wheals, reflects the difference in reaction of the patient to the 2 allergens. The value of *d_{rst}* can be 0, 1, 2, 3 or 4. If one allergen causes a strong skin response and the response to the other one is weak, then *d_{rst}* will be large (equals 2, 3 or 4). If the reaction to the 2 allergens is similar, either strong or weak, *d_{rst}* will be small (equals 0 or 1).

Two distances can be calculated for *K_{rs}* patients who were tested for their reaction to allergens *r* and *s*, but only *M_{rs}* of them responded positively to one or both allergens. The distances between allergens *r* and *s*, based on the responses of the patients that were tested for both of them, were calculated by the following equations:

$$d_{rs} = \frac{N}{M_{rs}} \sum_{t=1}^{M_{rs}} d_{rst} = \frac{N}{M_{rs}} \sum_{t=1}^{M_{rs}} |i_{rt} - i_{st}| \text{ and} \quad (1)$$

$$d_{rs}^* = \frac{N}{K_{rs}} \sum_{t=1}^{K_{rs}} d_{rst} = \frac{N}{K_{rs}} \sum_{t=1}^{K_{rs}} |i_{rt} - i_{st}|, \quad (1')$$

where *N* = 153 is the total number of patients tested.

The first measure (*d_{rs}*) does not take into account the cases in which a patient did not respond to either of the 2 allergens, while the second one (*d_{rs}^{*}*) includes such cases. The choice of the appropriate measure of dissimilarity depends on the group of allergens in question, as explained below.

The distances within all possible pairs of the 42 allergens were calculated. The distance within each pair of allergens to which the patients have responded in an identical way is zero. The distance between allergens to which the patients responded in a similar way was regarded as 'small', whereas the distance between allergens with different responses was regarded as 'large'.

The calculated distances among allergens were used as a basis for their rearrangement in an ordered list that reflects their overall similarities. The mathematical algorithm that we employed, in order to derive this list, is related to the arrangement of a list of items, while minimizing the path passing all of them, based only on the distances among them [12]. The analogy is planning a route of a traveling salesman, among a number of cities, in a way that will minimize his total traveling distance. This analogy gave the problem its famous name, the 'Travelling Salesman's Problem' (TSP).

In our case, the distance between every 2 allergens, which is a measure of their dissimilarity, was calculated on the basis of the patients responses to skin prick tests by equation (1). The solution of the TSP for these data resulted in an ordered list of allergens. Arrangement of allergens in a two-dimensional symmetric matrix according to this order yielded clusters of allergens to which the patients responded in a similar way. The average distance of each allergen to all others in its cluster was used to measure the similarity within the cluster.

The distances among allergens were also used for testing the discrimination among all pairs of identified clusters. When testing two clusters A and B consisting of *N_A* and *N_B* allergens, respectively, the mean distance among allergens within the clusters and between the clusters were calculated by the following equations:

$$D_A = \frac{2}{N_A(N_A-1)} \sum_{r,s \in A} d_{rs}, \quad (2)$$

$$D_B = \frac{2}{N_B(N_B-1)} \sum_{r,s \in B} d_{rs} \text{ and } D_{AB} = \frac{1}{N_A N_B} \sum_{r \in A, s \in B} d_{rs}$$

in which the sums are taken for all corresponding pairs of all allergens *r* and *s* (*r, s* ∈ *A* and *r, s* ∈ *B* indicate that both allergens belong to the same cluster, and *r* ∈ *A*, *s* ∈ *B* indicates that each allergen belongs to another cluster). The discrimination between clusters was tested by comparing the distances within clusters (*D_A* and *D_B*) with the mean distance between clusters *D_{AB}* using the two-sample *t* test [13].

This investigation was approved by the Institutional Ethical Review Boards of Tel Aviv University and the Kaplan Medical Center.

Results

The results illustrate the arrangement of 42 allergens on the basis of the responses of 153 patients. The degree of differences among allergens are indicated in the arranged

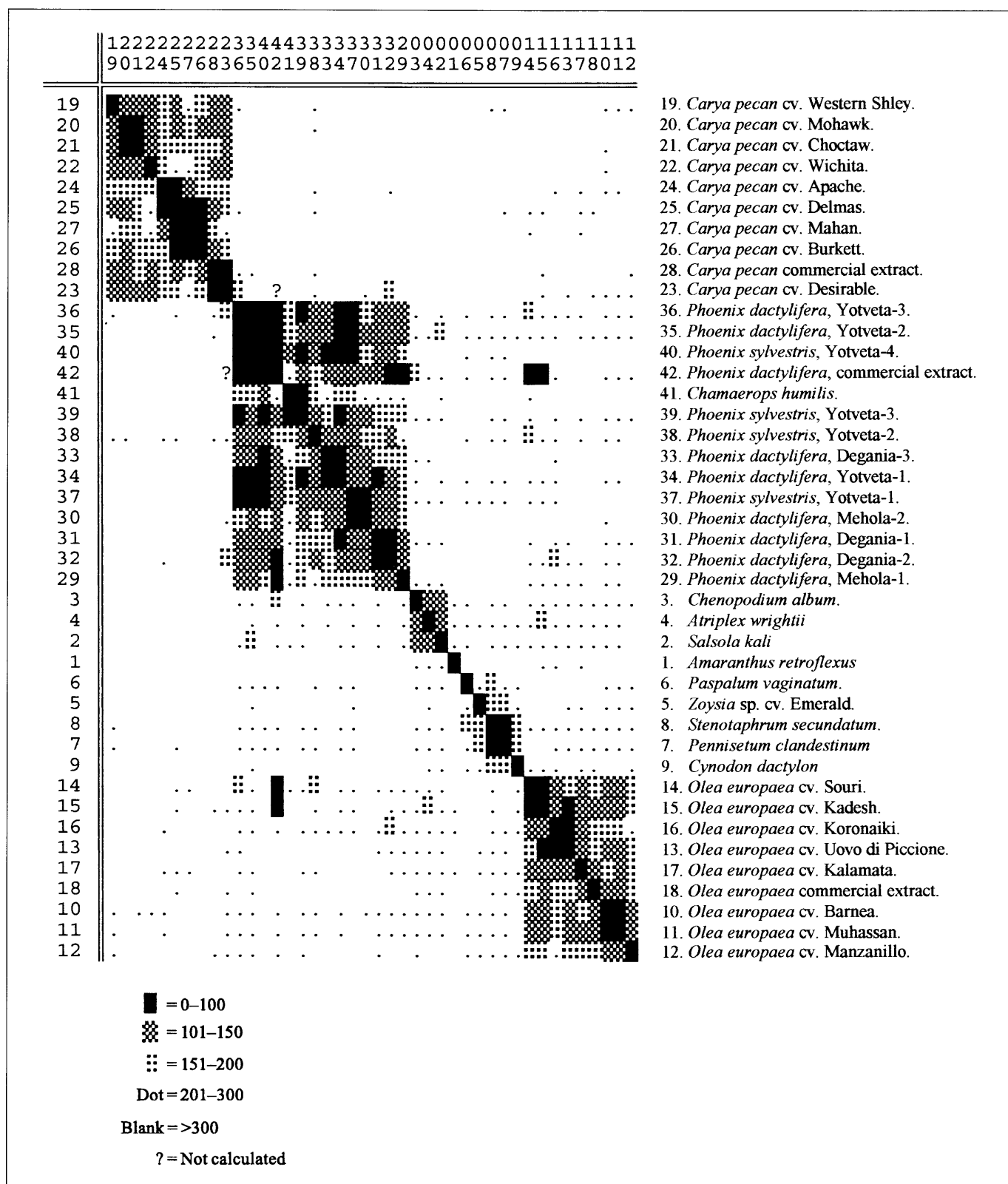


Fig. 1. Matrix of dissimilarities among all 42 allergens investigated. Distances were grouped and replaced by symbols.

symmetric matrix (fig. 1). The order of columns and lines of the matrix, which reflects the overall dissimilarity between allergen pairs, was determined by the matrix-structuring algorithm. The points along the diagonal of the matrix represent a 'zero difference', i.e. the dissimilarity of each allergen to itself. These form the axis of symmetry of the matrix. The actual values of dissimilarity were grouped and presented by symbols rather than by the actual numbers.

The exact structure of the matrix for any given group of allergens depends upon the measure of dissimilarity used, i.e. whether lack of reaction to 2 allergens was regarded as indicating similarity between them (equation 1') or not (equation 1). Equation (1) was used when the whole set of allergens was analyzed together. The alternative equation (1') was used only when a cluster of closely related allergens was tested. Therefore, we observed a certain difference between the analysis which was made for the full set of data and the one in which only a specific cluster was included. This is why the order of allergens within specific clusters (figs. 2, 4, 5) differs somewhat from that within the full set (fig. 1).

The obtained structure data have resulted in the formation of five clusters of allergens (fig. 1). Distances within clusters and between clusters were calculated by equations (2). The two-sample t test confirmed that distances within clusters were significantly smaller than distances between each pair of clusters at the $\alpha = 0.005$ level (table 1).

The pollen extracts of the pecan cultivars (allergens No. 19–28) form a distinct cluster. Nevertheless, even this cluster of pecans is not fully homogeneous. When this cluster is analyzed separately in greater detail, a subcluster, comprised of cultivars No. 24–28 (fig. 2a) can be distinguished. The allergens included in this cluster (with an average distance between them of 36.4) are more homogeneous than those of the entire cluster of pecans (average distance 113.8). The allergens of the other pecan cultivars (No. 19–23) appear to be less uniform. When the average dissimilarities of pecan allergens are calculated separately (fig. 2b), the commercial pecan extract (allergen No. 28) is located at the center of the cluster. This allergen shows a similar difference from nearly all other pecan extracts and has the smallest average measure of dissimilarity (74.7). Thus, the allergen of this source constitutes a fair representative of the eight pecan cultivars which were analyzed. The remaining 32 allergens are clearly detached from the pecans. Nonetheless, they are not homogeneous, and one can distinguish among them four distinct clusters of allergens (fig. 1). These include one cluster of palms (No. 29–42), one cluster which is constituted of three chenopods (No. 2–4) and one amaranth, one cluster of grasses (No. 5–9) and

Table 1. Mean distances among allergens within each of the five clusters identified in figure 1 (C1–C5, arranged along the diagonal of the table) and mean distances among allergens included in different clusters

C1	C2	C3	C4	C5	
149.6	368.3 ¹	373.4 ¹	370.0 ¹	319.0 ¹	C1
	132.7	307.5 ¹	313.7 ¹	287.4 ¹	C2
		181.2	298.2 ¹	261.3 ¹	C3
			193.1	281.7 ¹	C4
				137.4	C5

C1 = pecan cultivars (allergens 19–28); C2 = palms (allergens 29–42); C3 = chenopods and amaranth (allergens 1–4); C4 = grasses (allergens 5–9); C5 = olive cultivars (allergens 10–18).

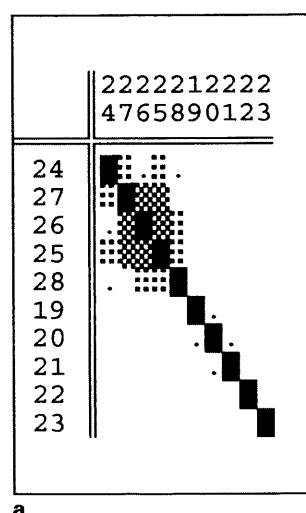
¹ Mean distances between clusters are significantly larger (two-sample t test at $\alpha = 0.005$) than those within clusters.

one cluster that includes all the tested olive cultivars (No. 10–18). Variability was relatively small among the various pollen sources of the palms (fig. 1). Nevertheless, the commercial extract of date palm pollen (No. 42) has a rather high level of dissimilarity (fig. 3) and is not a good representative of the tested palms.

Another cluster is formed by the pollen extracts of the three members of the Chenopodiaceae (*Salsola kali*, No. 2; *Chenopodium album*, No. 3, and *Atriplex wrightii*, No. 4) and the single representative of the Amaranthaceae, i.e. *Amaranthus retroflexus*, No. 1 (fig. 4). Within this cluster, the three chenopods exhibit a lower average dissimilarity, setting the *Amaranthus* aside. Thus, in spite of the commonly accepted cross-reactivity between the pollen allergens of the constituents of those two families [14], a significant dissimilarity between them is revealed by the presented matrix analysis.

Three of the five grasses investigated (No. 6–8) are closely related and form a conspicuous subcluster (fig. 4). Only *Zoysia* cv. 'Emerald' (No. 5) and *Cynodon* (No. 9), which are systematically distinct, exhibit some different allergenic traits. Thus, such a detailed analysis proves that dissimilar allergens can be found even among the highly cross-reactive grasses [15].

Olives (No. 10–18) form a distinct though variable cluster. Two relatively different subclusters can be observed when the cluster of olives is analyzed separately (fig. 5): allergens 14–18 and allergens 10–13. Allergens 12 and 15 are distinct among the tested olives, having the largest average dissimilarity (fig. 5b). The commercial extract of olive pollen (No. 18) is located in the center, i.e. between the two



a

■ = 0–25
 ▤ = 26–35
 ▦ = 36–45
 · = 46–55
 Blank = >55

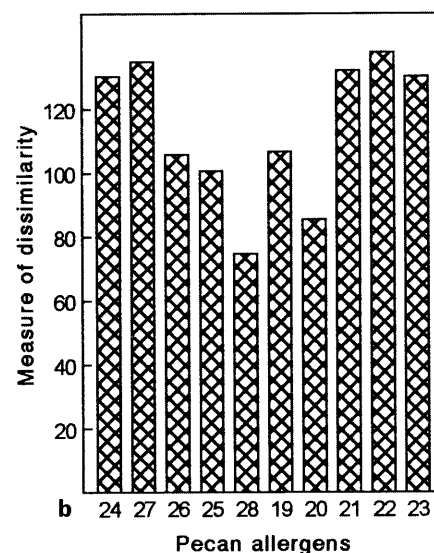


Fig. 2. Dissimilarities among the pecan allergens. **a** Matrix of distances. Distances were grouped and replaced by symbols. **b** Average dissimilarities within the cluster of pecan allergens ordinated by the matrix-structuring technique (see fig. 1 for identification of allergens).

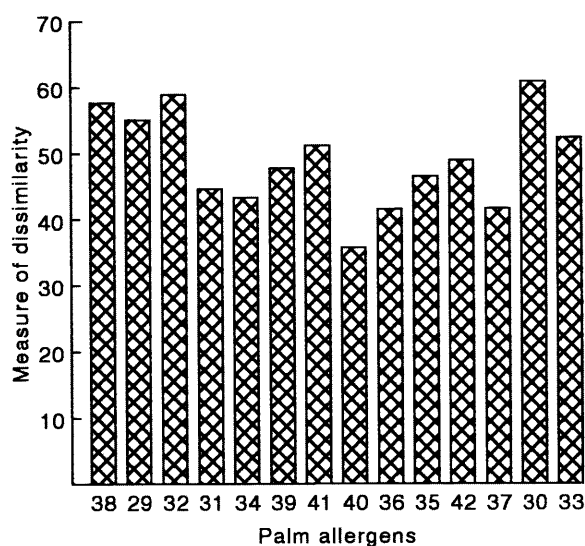
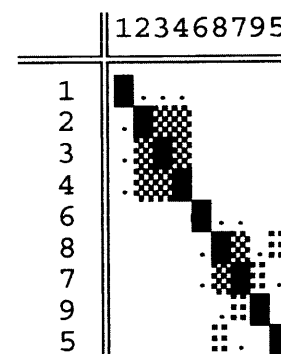


Fig. 3. Average dissimilarities within the cluster of palm allergens (*Phoenix* and *Chamaerops*) ordinated by the matrix-structuring technique (see fig. 1 for identification of allergens).



■ = 0–35
 ▤ = 36–50
 ▦ = 51–70
 · = 71–100
 Blank = >100

Fig. 4. Matrix of dissimilarities among the allergens of the five grasses, three chenopods and the amaranth investigated. Distances were grouped and replaced by symbols (see fig. 1 for identification of the allergens).

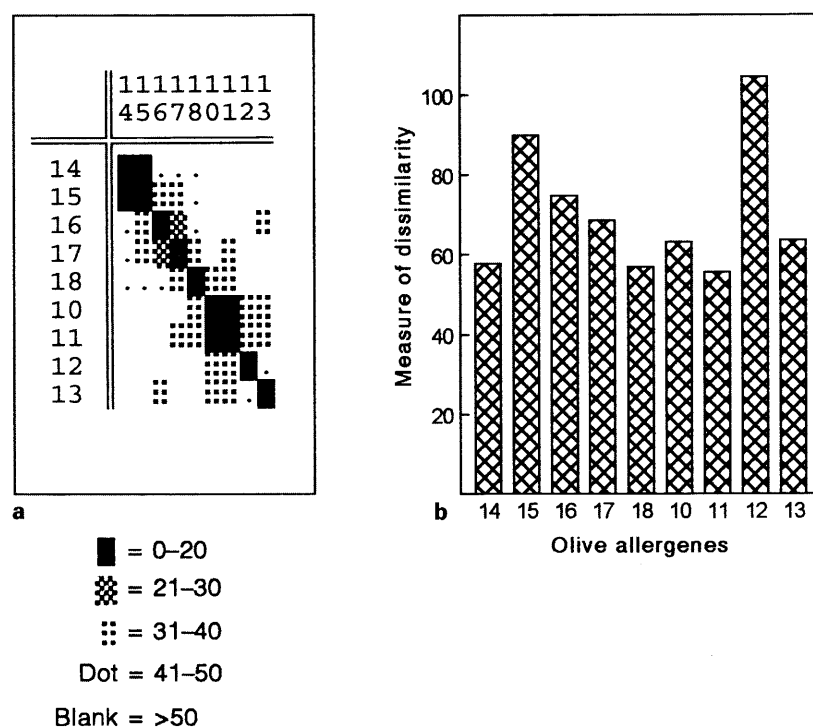


Fig. 5. Dissimilarities among the olive allergens investigated. **a** Matrix of distances. Distances were grouped and replaced by symbols. **b** Average dissimilarities within the cluster of olive allergens ordinated by the matrix-structuring technique (see fig. 1 for identification of the allergens).

mentioned subclusters (fig. 5a), with the smallest value of average dissimilarity (fig. 5b).

It is interesting to note that a certain degree of similarity was found between the responses to two olive cultivars ('Souri' No. 14 and 'Kadesh' No. 15) and to one of the palms (*P. dactylifera*, No. 42) (fig. 1). However, because of the relatively small numbers of patients that were tested for reactivity to both groups of allergens, such an overlapping of allergens should be further investigated.

Discussion

The general idea of the presented data-processing technique is to arrange the allergens in a certain order according to the patient's response. The structured form of the data enables the investigator to view the allergens from a new epidemiological standpoint and to reveal possible concealed relationships. Apparently, the presented method demonstrates its effectiveness in the analysis of skin test responses of a large number of patients. It is proven useful for the ranking of similarity among closely related allergens or dissimilarity among distant ones [10]. This method of analysis is not intended to replace the precise analysis of antigen

similarity based on RAST, ELISA or on other similar immunological methods. Nevertheless, it gives a cheap, easy and convenient way for fast procession of information regarding responses of a large population of patients to a large number of allergens.

Moreover, analysis of skin test responses, which are easy to obtain for a large number of patients, can point out the existence of similarities among distinct allergens and reveal certain degrees of cross-reactivity. The similarity between the results of our present study and those of previous ones [10], based on different groups of patients, is a good proof for the consistency and usefulness of this method. Thus, the presented solution yields sensible answers to the presented requirements.

Most allergies are caused by pollen of heterogeneous plant species. How can one distinguish between relevant and irrelevant test materials? Do the commercially available pollen extracts, whose sources are far and diverse [16–18], give a fair representation of the local allergens? Is it feasible to standardize the pollen extracts of hundreds of different cultivars of olives, date palms and pecans? The answer to such a question is dual: 'yes' in very few cases (mostly for the luxuriant clinics of the 'rich' countries), but 'no' for all others. When an extract has a small average measure of dis-

similarity compared with other extracts, it should be considered as a good representative. However, in cases where a purchased extract is considerably dissimilar to the local allergens, or when it represents only a subgroup, one should pursue the use of a more specific extract, preferably one that is made from locally collected pollen. In fact, results of the present study enable us to assert that the commercial pollen extracts of some species are acceptable representatives of the local allergens. In contrast, the suitability of the commercial pollen extracts of the palms and olives is questionable and should be further evaluated. In such cases, a double check, using two different sources, is recommended [19].

'Full cross-reactivity' between the allergens of closely related plants is frequently assumed when similar responses of patients are recorded [6, 11, 17, 20, 21]. Are the pollen of various sources really equally cross-reactive? The answer is probably 'no' [16–18]. Even when standardized allergens are used, only some people respond to a few of the 'cross-reactive' allergens, whereas other patients do not [19, 22]. It is interesting to note that 4 out of the 5 allergens included in the first cluster (cv. Souri, Kadesh, Koronaiki and Kalamata) exhibited a marked Ole e 1 band, whereas 3 of 4 allergens of the second cluster (cv. Barnea, Manzanillo and Uovo di

Piccione) do not have this protein [22]. Such dissimilarity is further confirmed by the presented analysis, where the existence of minor nuances in allergenicity of the tested amaranth and the three chenopods are shown [14]. Thus, for precise identification of the sources of allergy, and certainly for immunotherapy, distinction should be made between such subgroups, and the responses to chenopod pollen should be separated from those to amaranths [10].

The same is true for the analysis of various grasses. Turf grasses constitute a major source of allergenic pollen all over the world. Hence, the capability to detect turf grass species of low allergenicity, e.g. Zoysia cv. 'Emerald', serves as an excellent example for the practical aspects that are embedded in the suggested TSP solution.

Acknowledgements

The support to E. Kosman by the Ministry of Immigrant Absorption and of the Gileady Programme is gratefully acknowledged. We are much indebted to Drs. Z. Dolev, M. Finelt, C. Geller-Bernstein, A. Goldberg, S. Kivity, I. Panasoff, D. Nusem and E. Shahar for participating in this investigation and for performing the skin tests. Special thanks are due to Dr. N. Keynan for her indispensable help.

References

- Middleton E, Reed CE, Ellis EF (eds): Allergy: Principles and Practice. Mosby, Toronto, 1983, pp 43–74, 271–294.
- Roitt I, Brostoff J, Male D (eds): Immunology, ed 2. London, Gower, 1989.
- Berzofsky JA, Schechter AN: The concept of cross reactivity and specificity in immunology. Mol Immunol 1981;18:751–763.
- Bousquet J, Guerin B, Hewitt B, Lim S, Michel FB: Allergy in the Mediterranean area. III. Cross reactivity among Oleaceae pollens. Clin Allergy 1985;15:439–448.
- Bousquet J, Hewitt B, Guerin B, Dhivert H, Michel FB: Allergy in the Mediterranean area. II. Cross reactivity among Urticaceae pollen (*Parietaria* and *Urtica*). Clin Allergy 1986;16:57–64.
- Corbi AL, Pelaez A, Errigo E, Carreira J: Cross reactivity between *Parietaria judaica* and *Parietaria officinalis*. Ann Allergy 1985;54:142–147.
- Weber R: Cross reactivity among pollens. Ann Allergy 1981;46:208–215.
- Waisel Y, Mekori J: Environmental pollution with pollen and its effect on the frequency of allergic rhinitis and asthma among agricultural workers. Research Report. The Committee for Research & Prevention in Occupational Safety and Health. Jerusalem, Ministry of Labour and Social Affairs, Israel 1992.
- D'Amato G, Mullins J, Nolard N, Spiekma FTHM, Wachter R: City spore concentrations in the European Economic Community (EEC). VII. Oleaceae (*Fraxinus*, *Ligustrum*, *Olea*). Clin Allergy 1988;18:541–547.
- Kosman EG, Keynan N, Eshel A, Waisel Y: Clustering of allergenic pollen on the basis of skin responses of atopic patients by matrix analysis. Allergy 1994;49:502–507.
- Keynan N, Geller-Bernstein C, Waisel Y, Bejerano A, Shomer-Ilan A, Tamir R: Positive skin tests to pollen extracts of four species of *Pistacia* in Israel. Clin Allergy 1987;17:243–249.
- Kosman EG: Matrix structuring. Soviet Math Dokl 1990;41:152–155.
- Statistix 3.5. An Interactive Statistical Analysis Program for the IBM PC. St Paul, Analytical Software, 1991.
- Lombardero M, Duffort O, Selles JG, Hernandez J, Carreira J: Cross-reactivity among Chenopodiaceae and Amaranthaceae. Ann Allergy 1985;54:430–436.
- Esch RE, Klapper DG: Isolation and characterization of a major cross-reactive grass group I allergenic determinant. Mol Immunol 1989;26:557–561.
- Malik P, Singh AB, Gongal SV, Babu CR: Comparison of antigenic and allergenic components of *Holoptela integrifolia* pollen collected from different source materials. Allergy 1991;46:284–291.
- Maasch HJ, Hauck PR, Oliver JD, Geissler W, Wahl R, Winter HG: Allergenic activity of short ragweed (*Ambrosia elatior*) from different years and/or suppliers: Criteria for the selection of an inhouse allergen reference preparation. Ann Allergy 1987;58:429–434.
- Wheeler AW, Hickman BE, Fox B: Heterogeneity of a major allergen from olive (*Olea europaea*) pollen. Mol Immunol 1990;27:631–636.
- Waisel Y, Geller-Bernstein C: Reliability of olive pollen extracts for skin prick tests. J Allergy Clin Immunol 1996;98:715–716.
- Waisel Y, Geller-Bernstein C, Shomer-Ilan A, Tamir R, Keynan N: Allergy to *Pistacia*; in Basomba A, Sastre J (eds): ECACI Madrid. Bologna, Mondozi, 1995, pp 665–669.
- Wheeler AW: Hypersensitivity to the allergens of the pollen from the olive tree (*Olea europaea*). Clin Exp Allergy 1992;22:1052–1057.
- Waisel Y, Geller-Bernstein C, Keynan N, Arad G: Antigenicity of the pollen proteins of various cultivars of *Olea europaea*. Allergy 1996;51:819–825.