Food Chemistry 346 (2021) 128929

Contents lists available at ScienceDirect

Food Chemistry

journal homepage: www.elsevier.com/locate/foodchem

Analytical Methods

Characterization of herbal teas containing lime flowers – *Tiliae flos* by HPTLC method with chemometric analysis

Natalia Melnyk [a,b], Karolina A. Pawłowska [a], Maria Ziaja [c], Wojciech Wojnowski [d], Oleh Koshovyi [b], Sebastian Granica [a,*], Agnieszka Bazylko [a]

[a] Department of Pharmacognosy and Molecular Basis of Phytotherapy, Centre for Preclinical Studies, Faculty of Pharmacy, Medical University of Warsaw, ul. Banacha 1, 02-097 Warsaw, Poland
[b] National University of Pharmacy, 53 Pushkinska Str., Kharkiv, 61002, Ukraine
[c] Institute of Physical Culture Studies, Rzeszów University, Cicha 2a, 35-326 Rzeszów, Poland
[d] Department of Analytical Chemistry, Faculty of Chemistry, Gdańsk University of Technology, Gdańsk 80-233, Poland

ARTICLE INFO

Keywords:
Tiliae flos
Linden flowers
HPTLC
Chemometric analysis
Chemical analysis

ABSTRACT

Linden trees are a source of food products called lime flowers (*Tiliae flos*), traditionally used in the form of infusion for the treatment of feverish colds and coughs. Lime flowers should include flowers of *Tilia cordata* Mill, *T.x europaea* L., and *T. platyphyllos* Scop. or a mixture of these. The aim of current research was to establish a fast, sensitive HPTLC (high-performance thin-layer chromatography) method that would allow the differentiation of material obtained from five species of lime occurring in Europe. The fingerprints for distinguishing these species were established, as well as a key for identification based on a visual evaluation of chromatograms. The results obtained were also subjected to chemometric analyses. It was shown that each species contains characteristic compounds i.e. linarin that can be used for their identification. The method developed can, in theory, be introduced for the quality control or authentication of linden flowers on the European market.

1. Introduction

Tilia sp. (Tiliaceae) are deciduous trees and bushes native throughout the temperate regions of the Northern Hemisphere (Pigott, 2012). Linden trees are a source of food products and pharmacopeial plant material called *Tiliae flos* (lime flowers), which consist of pedicel bearing 3–6 yellowish-white, five-petalled fragrant flowers connected to an oblong bract (Pigott, 2012). *Tiliae flos* has a monograph in European Pharmacopeia 9.0 (2016), and also its description was published by EMA (European Medicines Agency) as a herbal traditional medicine (2011, 2012). In traditional medicine lime flowers are used in the form of an infusion for the treatment of feverish colds, catarrh, coughs, and influenza (Heinrich et al., 2018). Extracts from *Tiliae flos* are also used as diaphoretics, antispasmodics, and expectorants (Evans et al., 2009). According to the pharmacopeial description, lime flowers should consist of the dried inflorescence of *Tilia cordata* Mill, *T.* × *europaea* L., and *T. platyphyllos* Scop., or a mixture of these. The monography in Eur. Ph. 9.0 mentions that the raw plant material must not contain more than 2% of flowers of *T. americana* L. or *T. tomentosa* Moench that can be easily accessed in Europe. According to the pharmacopeia the evaluation of so-called foreign matter in the test section of the monography is based on macro- and microscopic assessment. The contamination with *T. americana* or *tomentosa* is confirmed by the presence of hexamerous flowers in the raw material (2016). It is not clear whether the chemical composition of flowers depending on the species from which they were obtained is crucial for their therapeutic uses. Some studies show that no significant differences in the phenolics profile of extracts from common European species have been observed by UHPLC (ultra-high-performance liquid chromatography) analysis (Karioti et al., 2014). In some regions of the world, all five *Tilia* species are used as medicinal plant material. It is worth mentioning that EMA also issued a final assessment report on *T. tomentosa* in 2012 (2011, 2011).

In the literature there is no fast or cheap method based on TLC (thin-layer chromatography) for the differentiation of extracts prepared from flowers collected from different *Tilia* species. A simple TLC method for the authentication of the plant material focused on the evaluation of the phenolics profile is described in the pharmacopeial monography. However, the monography does not mention if any variability in the

* Corresponding author.
E-mail addresses: mziaja@ur.edu.pl (M. Ziaja), wojciech.wojnowski@pg.edu.pl (W. Wojnowski), sgranica@wum.edu.pl, sgranica@gmail.com (S. Granica), agnieszka.bazylko@wum.edu.pl (A. Bazylko).

https://doi.org/10.1016/j.foodchem.2020.128929
Received 19 August 2020; Received in revised form 10 November 2020; Accepted 20 December 2020
Available online 27 December 2020

chemical profile is possible in the case of flowers originating from different linden tree species.

The aim of the current research was to establish a fast yet sensitive HPTLC (high-performance thin-layer chromatography) method that would allow the differentiation of flowers obtained from five *Tilia* species, namely *T. cordata*, *T. platyphyllos*, *T.* × *europaea*, *T. americana*, and *T. tomentosa*, occurring in Europe. Taking into consideration that the commercial supplies of this medicinal plant material available on the market in Europe come from China, the Balkans, Turkey, and Hungary (Evans et al., 2009), and the morphological identification of their source is often not possible due to the lack of proper documentation, a fast yet reliable method for the authentication of their source species and contaminants is needed. Moreover, the current research aimed at the development of a methodology that would allow for the identification of lime flowers contained in herbal teas and other products present on the market.

2. Materials and methods

2.1. Solvents and reagents used in the study

Tetrahydrofuran for HPTLC, dichloromethane, acetic acid, sulphuric acid, isopropanol, methanol, isopropyl ester, ethyl acetate, toluene, chloroform were purchased from POCh (Gliwice, Poland); Naturstoffreagenz A, vanillin, and formic acid were purchased from Merck (Darmstadt, Germany). Distilled water was purified with the Millipore Simplicity UV system (Bedford, MA, USA).

Chemical standards (rutin, kaempferol, linarin, isoquercetin, astragalin, avicularin, *trans*-tiliroside) were isolated from *Tilia cordata* flowers in the Department of Pharmacognosy and Molecular Basis of Phytotherapy of Medical University of Warsaw (Ziaja et al., 2020). Hyperoside and caffeic acid were purchased from Sigma Aldrich (Darmstadt, Germany). A list of standards used in the study is given in Table S1 (Supplementary materials). All standards were of HPLC grade. The purity was over 95% by HPLC-DAD at 254 nm.

2.2. Plant material used for the study

2.2.1. Tilia samples harvesting and handling

Tiliae flos came from 5 species: *T. tomentosa* Moench (TT), *T.* × *europaea* L. (TE), *T. americana* L. (TA), *T. cordata* Mill. (TC), and *T. platyphyllos* Scop. (TP), collected in Poland and other European countries, between 2014 and 2016. The plant material was authenticated by Maria Ziaja according to their morphological characteristics (Rutkowski, 2014). Samples were harvested from several locations in Poland and one location in Germany and Italy, respectively. From over 250 samples 35 representative samples were chosen (at least 4 per species). Samples came from different years in order to eliminate intra-seasonal variability in their chemical profiles and from different locations in order to reduce the influence of ecological conditions. Flowers were harvested in full bloom and air-dried. They were then stored in darkness at room temperature in a dry place for further use. List of samples used in the study together with voucher numbers and GPS coordinates of collection sites are listed in Table S2 (Supplementary materials).

2.2.2. Commercial herbal teas containing lime flowers used in the current study

Herbal teas containing *Tiliae flos* as the major or only ingredient were purchased in community pharmacies or in other herbal stores around Europe. Samples were described according to the manufacturer's declaration. Detailed data about samples (TS1–TS30) used for experiments are given in Table S4 (Supplementary materials).

2.3. Preparation of extracts from investigated samples for HPTLC analysis

One gram of ground raw material was extracted with 20 mL of methanol:water (1:1, *v/v*), at 40 °C using a sonication bath (15 min/40 °C). Next, the sample was centrifuged at 9503×*g* for 3 min. Supernatants were filtrated through cotton wool and paper filters to evaporation flasks. The plant material was extracted once more with 10 mL of methanol:water (1:1, *v/v*). The obtained extracts were combined. Extracts were evaporated under reduced pressure with a LABORANTA 4000 WB evaporator (<45 °C, Heidolph, Schwabach, Germany), and their residue was suspended in 10 mL of water. Samples were freeze-dried with Cryodos apparatus (Telstar, Terrassa, Spain). Dry extracts were immediately transferred to glass vials, sealed, and stored at 4 °C for further use.

2.4. Preparation of solutions from standards and extracts for HPTLC application

Sample solutions were prepared by dissolving extracts in methanol: water (8:2, *v/v*) mixture (20 mg/mL). Standard solutions were prepared by dissolving accurately weighed standard substances in pure methanol in order to obtain stock solutions at a concentration of 1 mg/mL (kaempferol 3-*O*-glucoside-7-*O*-rhamnoside – GRK, isoquercetin – IQ, rutin – R, avicularin – AV, linarin – L, hyperoside – H and caffeic acid – CA), or 0.5 mg/mL (astragalin – AS and *trans*-tiliroside – TZ).

2.5. HPTLC apparatus and chromatographic conditions

Chromatographic analysis was carried out on a TLC system from Camag (Muttenz, Switzerland), which consisted of a Linomat 5 equipped with 100 μL Hamilton syringe, Automatic Developing Chamber 2 (ADC2), Scanner 4, Automatic Derivatizer, Visualizer 2, and a Plate Heater III. HPTLC plates (20 × 10 cm) (Merck, Darmstadt, Germany) precoated with silica gel 60 F254 were used as a stationary phase. The plates were prewashed with methanol, dried for 10 min at room temperature, and activated at 80 °C for 30 min. Samples of extracts (20 mg/mL) and samples of standards solutions (1 mg/mL) were applied (5 μL) on plates as 8.5 or 9 mm bands. The other operating parameters like the distance between tracks, distance from the plate side edge, and distance from the bottom of the plate were set at 11, 17, and 10 mm, respectively. The plates were developed in ADC2 using tetrahydrofuran:dichloromethane:formic acid:acetic acid:water (9:9:4:2:3, *v/v/v/v/v*) as the mobile phase. Parameters of the development were as follows: humidity control – 2 min, tank saturation – 15 min, plate preconditioning – 5 min, temperature 22 ± 1 °C, humidity 50 ± 2%, migration distance 85 mm, plate drying – 10 min.

Spots were visualized under UV light at 254 and 366 nm, and at 366 nm after spraying with 1.0% methanolic diphenylboric acid-β-ethylamine ester (Naturstoffreagenz A, Roth, Karlsruhe, Germany) (NA). Densitograms were recorded at 254 nm before derivatization.

Plates were heated at 100–105 °C for 2 min and derivatized using Automatic Derivatizer (setting used – level 3, Camag). Sprayed plates were documented as white light and UV light at 254 and 366 nm. All the operating parameters were controlled by winCATS software 1.4.2 (Camag).

2.6. Chemometric analysis of obtained data

The multivariate statistical analysis was carried out in Python 3.7 using Orange v. 3.23, Scikit-learn v. 0.23, and LIBSVM v. 3.24 packages (Chang & Lin, 2007; Demšar et al., 2013; Pedregosa et al., 2011). The semi-quantitative data (peak areas at 254 nm) obtained for samples TA, TC, TE, TP, and TT were used as input and normalized. The relationship between features and the normalized attribute values was visualised using a heat map with clustering. In order to avoid overfitting, the most

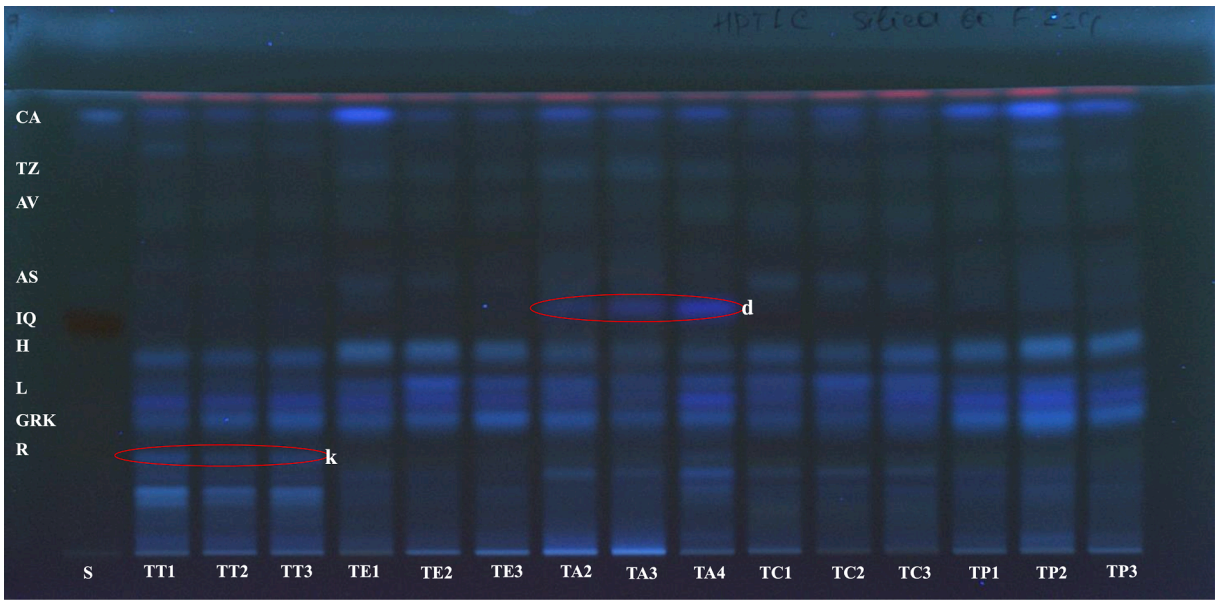


Fig. 1. HPTLC chromatogram at 366 nm of five *Tilia* species before derivatization with 1% NA (Naturstoffreagenz A).

relevant features were chosen based on the analysis of variance using the ReliefF algorithm, which is optimized for sensitivity to feature interactions, particularly with discrete features (Tabachnick & Fidell, 2006; Todorov, 2016). The relationship between objects (samples) was investigated based on Euclidean distances between the data points, using hierarchical clustering with Ward linkage and using linear discriminant analysis (LDA), coupled with principal component analysis (PCA). The TS samples were tentatively classified using the support vector machines (SVM) machine learning model with RBF kernel. The model was validated through 10-fold stratified random sampling, with 66% of the data set used for training, and the remaining 34% used for testing, and also through 10-fold stratified cross-validation. The areas for peaks detected at 254 nm in recorded densitograms were used as semi-quantitative data (Table S5, Supplementary materials).

3. Results and discussion

3.1. Optimization of HPTLC system – The choice of chromatographic conditions

In order to establish which group of phytochemicals contained in the analyzed extracts (procyanidins or phenolic acids/flavonoids) would be more suitable for the determination of the differences in the phytochemical profiles of the chosen *Tilia* species two derivatization reagents [Naturstoffreagenz A (Natural Products reagent) for phenolics and vanillin in 96% sulphuric acid for flavan-3-ols] were used for the visualization of spots after the plate development. Following these tests, the focus was on phenolic acids/flavonoids analysis using Naturstoffreagenz A as a derivatization reagent.

The main goal of the optimization part of this HPTLC research was to find a suitable mobile phase and obtain a good resolution of major compounds found in lime flower extracts. Silica gel was used as the most

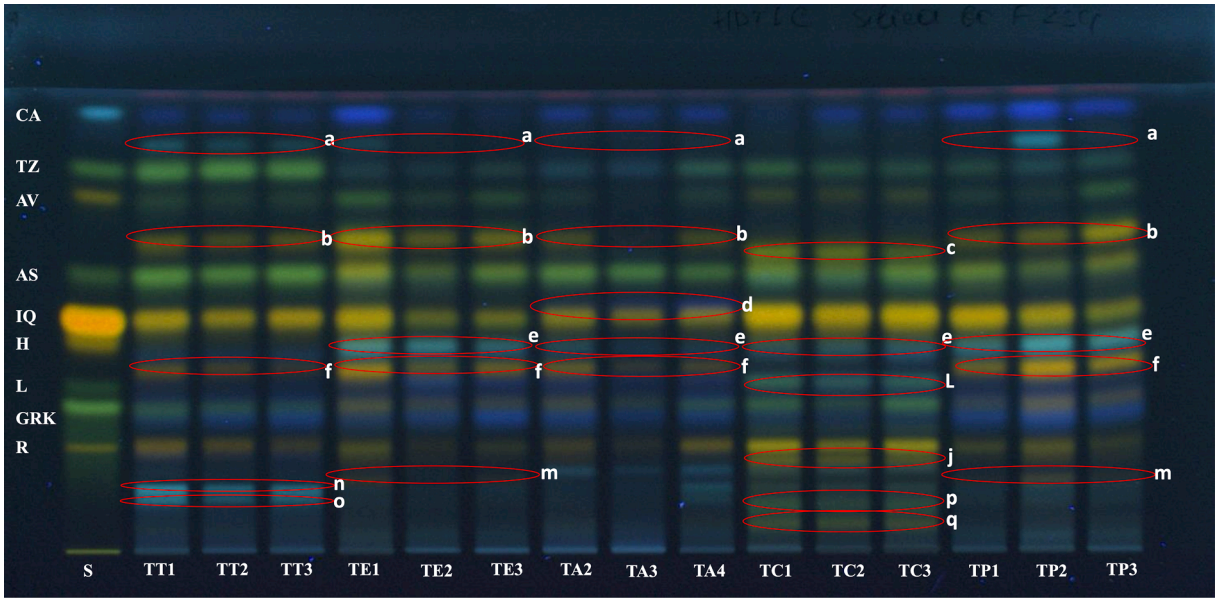


Fig. 2. Representative HPTLC chromatogram at 366 nm of five *Tilia* species after derivatization with 1% NA (Naturstoffreagenz A).

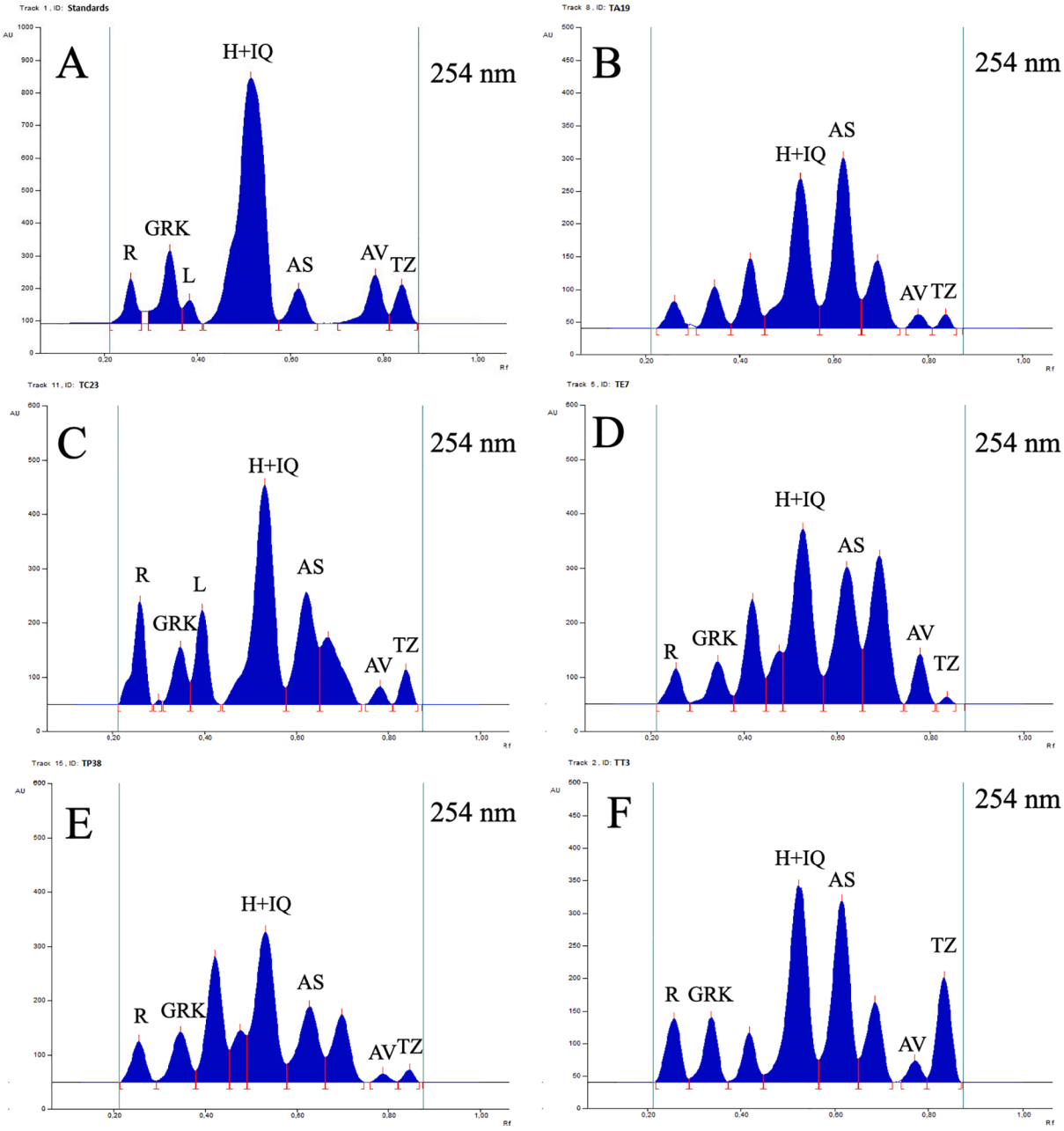


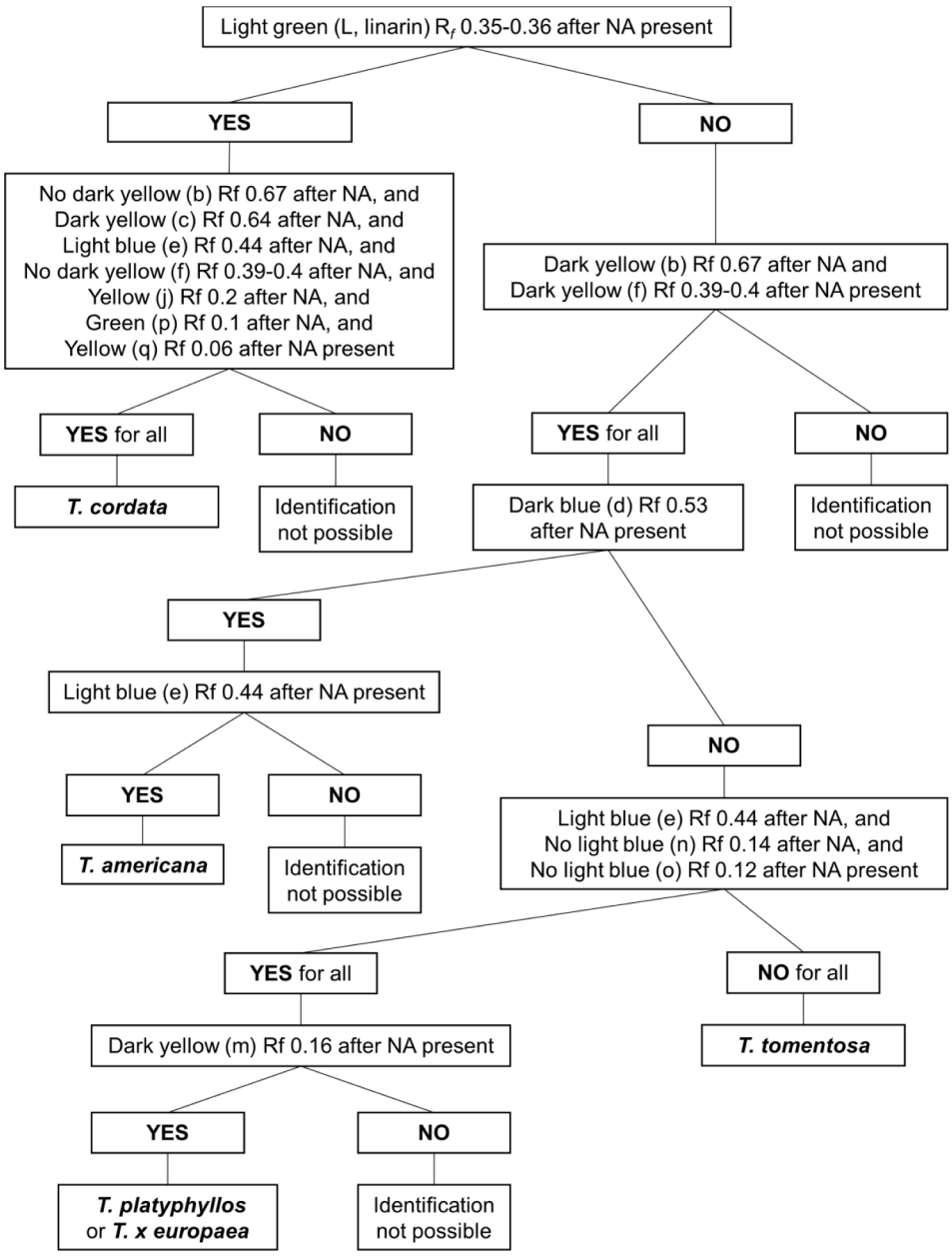
Fig. 3. Representative HPTLC densitograms recorded at 254 nm before spraying with 1% NA (Naturstoffreagenz A). A – standards mixture, B – *T. americana*, C – *T. cordata*, D – *T.* × *europaea*, E – *T. platyphyllos*, F – *T. tomentosa*, R (rutin), GRK (kaempferol 3-*O*-glucoside-7-*O*-rhamnoside), H (hyperoside), IQ (isoquercitrin), AS (astragalin), AV (avicularin), TZ (*trans*-tiliroside) - standards used in the study accordingly to Table S1.

convenient stationary phase for routine analysis. The optimization of the separation conditions was performed according to the Reich and Schibli procedure (Reich & Schibli, 2007).

After the first step of checking, where solvents from different selectivity groups according to Snyder's classification of solvents (for polar absorbents) (Snyder, 1971, 1978) were used, an interesting starting point for further analyzes was provided by the tests with ethyl acetate, tetrahydrofuran and dichloromethane. Because mobile phases with ethyl acetate are widely used for the TLC separation of flavonoids therefore, at the beginning mobile phases containing this solvent as a major solvent were checked. Unfortunately, sufficient separation of linarin (L) and kaempferol derivative (GRK) was not obtained. Finally, the best effects were obtained with a mixture of tetrahydrofuran: dichloromethane:formic acid:acetic acid:water (9:9:4:2:3, *v/v/v/v/v*) as the mobile phase, as it led to the best results for the resolution of compounds, the shape of spots, and showed satisfactory fingerprints for *Tilia* species.

During the optimization of resolution, TLC plates were replaced with HPTLC plates, which resulted in the obtaining of a narrow shape of separated zones and more visible resolution (Fig. S1, Supplementary materials).

It was observed that the heating of HPTLC plates had a significant impact on visualization. Heating after derivatization caused differences between kaempferol and quercetin derivatives to be less pronounced, while heating before spraying led to the differentiation of quercetin glycosides as yellow spots, and kaempferol derivatives as green-yellow spots. For the above reason, conditions of heating using Naturstoffreagenz A as a derivatization reagent were as follows: heating on the Plate Heater III for 2 min and derivatization right after heating when the plate was still warm.



3.2. Description of differences in HPTLC profiles of investigated Tilia species

The characteristic fingerprints of investigated *Tilia* species obtained with the developed HPTLC method are shown in Figs. 1 and 2. It was possible to observe spots (both at 366 nm before spraying and at 366 nm after spraying) that seem to be characteristic for investigated species. Such specific spots were identified for *Tilia tomentosa:* light blue (a) below caffeic acid R_f 0.87 (Figs. 1 and 2), dark yellow (b) between avicularin and astragalin R_f 0.67 (Fig. 2), dark yellow (f) above linarin R_f 0.39–0.4, and absence of light blue (e) below isoquercetin R_f 0.44 (Fig. 2). *Tilia americana* can be identified by the presence of four zones: dark yellow (b) between avicularin and astragalin R_f 0.67(Fig. 3), dark blue (d) above isoquercetin R_f 0.53 (Figs. 1 and 2), light blue (e) below isoquercetin R_f 0.44, and dark yellow (f) above linarin R_f 0.39–0.4 (Fig. 2). *Tilia platyphyllos* and *Tilia europaea* had common fingerprints and their difference from other species is the presence of a dark yellow spot (b) between avicularin and astragalin R_f 0.67 (Fig. 2), light blue (e) below isoquercetin R_f 0.44 (Fig. 3), a dark yellow spot (f) above linarin R_f 0.39–0.4 (Fig. 2), and a dark yellow spot (m) below rutin R_f 0.16 (Fig. 2). This can be explained by the fact that *T. europaea* is a hybrid genus of *T. platyphyllos* and *T. cordata*. A large number of specific zones were identified for *Tilia cordata*: dark yellow (c) right above astragalin R_f 0.64 (Fig. 2), light blue (e) below isoquercetin R_f 0.44 (Fig. 2), linarin (L) R_f 0.35–0.36 (Fig. 2), yellow (j) R_f 0.2 (Fig. 2), green (p) R_f 0.1(Fig. 2) and yellow (q) R_f 0.06 below rutin (Fig. 2). The absence of light blue (a) below caffeic acid R_f 0.87 (Figs. 1 and 2), dark yellow (b) between avicularin and astragalin R_f 0.67 (Fig. 2), dark yellow (f) above linarin R_f 0.39–0.4 (Fig. 2) is used as a marker, as these zones are present in all species except *T. cordata.*

The differences are summarized in Table S3 (Supplementary materials) and are presented in a semi-quantitative manner.

Table 1
Authentication of herbal products based on visual evaluation and chemometric approach.

Sample ID	Declared species by manufacturer	Authentication based on visual evaluation	Authentication based on chemometric analysis
TS 1	*T. cordata*	*T. cordata* + *T. platyphyllos*/ *T. europaea*	*T. cordata*
TS 2	*T. platyphyllos*	*T. americana*	*T. cordata*
TS 3	*T. platyphyllos*	*T. americana*	*T. cordata*
TS 4	Not given	*T. cordata* + *T. americana*	*T. americana*
TS 5	*T. platyphyllos*	*T. americana*	*T. platyphyllos*
TS 6	*T. europaea*	*T. cordata*	*T. cordata*
TS 7	Not given	*T. platyphyllos*/ *T. europaea* + *T. cordata*	*T. cordata*
TS 8	Not given	*T. americana* + *T. cordata*	*T. cordata*
TS 9	Not given	*T. cordata* + *T. platyphyllos*/ *T. europaea*	*T. cordata*
TS 10	Not given	*T. platyphyllos*/ *T. europaea*	*T. cordata*
TS 11	Not given	*T. cordata* + *T. platyphyllos*/ *T. europaea*	*T. americana*
TS 12	Not given	*T. cordata* + *T. platyphyllos*/ *T. europaea*	*T. americana*
TS 13	*T. cordata*	*T. cordata* + *T. platyphyllos*/ *T. europaea*	*T. cordata*
TS 14	*T. europaea*	*T. americana*	*T. platyphyllos*
TS 15	Not given	*T. cordata*	*T. cordata*
TS 16	Not given	*T. platyphyllos*/ *T. europaea* + *T. cordata*	*T. platyphyllos*
TS 17	Not given	*T. tomentosa*	*T. cordata*
TS 18	Not given	*T. cordata* + *T. platyphyllos*/ *T. europaea*	*T. platyphyllos*
TS 19	Not given	*T. cordata* + *T. platyphyllos*/ *T. europaea*	*T. cordata*
TS 20	Not given	*T. platyphyllos*/ *T. europaea*	*T. cordata*
TS 21	Not given	*T. platyphyllos*/ *T. europaea*	*T. cordata*
TS 22	Not given	*T. platyphyllos*/ *T. europaea* + *T. cordata*	*T. cordata*
TS 23	Not given	*T. cordata* + *T. platyphyllos*/ *T. europaea*	*T. cordata*
TS 24	*T. cordata*	*T. tomentosa*	*T. cordata*
TS 25	Not given	*T. cordata*	*T. cordata*
TS 26	Not given	*T. cordata*	*T. cordata*
TS 27	Not given	*T. cordata*	*T. cordata*
TS 28	Not given	*T. cordata*	*T. cordata*
TS 29	Not given	*T. cordata* + *T. platyphyllos*/ *T. europaea*	*T. cordata*
TS 30	Not given	*T. cordata*	*T. cordata*

3.3. Development of the key for HPTLC identification of chosen species based on visual evaluation of HPTLC results

Based on the obtained results for the species comparison, a key for the phytochemical identification of five species was developed. The key is presented in the form of a decision tree. Following each branch, and answering questions, the identification of one of five investigated species is possible. The text version of the key is provided in the Supplementary materials.

Hierarchical key based on visual observation of HPTLC results preneted as a decision tree (please see Fig. 2).

The identification of *Tilia* species turned out to be possible based on the observation of HPTLC chromatogram, after spraying, and heating at 366 nm.

3.4. Application of the visual key for the identification of species contained in lime products

The observation of HPTLC data obtained for samples of thirty herbal products with different status, together with the phytochemical key developed in Section 3.3 (Figs. S2–S5, Supplementary materials), was performed. Where possible, each herbal product was classified into one of five investigated species *T. cordata*, *T. platyphyllos*, *T. americana*, *T. tomentosa*, *T.* × *europaea*, or as a mixture of them.

According to the key developed for the identification of five species, analyses of the thirty lime products purchased were carried out. The pharmacopeial monograph of *Tiliae flos* allows three species as a source of medicinal plant material to be categorized as an OTC drug. A mixture of three species (TP, TC, and TE) in any ratio is also possible. It also excludes *T. americana* and *T. tomentosa* as a valid source.

The HPTLC analyses of all five species (Figs. 1 and 2) showed that an extract from each of them can be easily identified based on the occurrence or absence of characteristic spots. Thus, visual observation should enable the identification of mixtures of linden flowers in investigated herbal products. The current investigation proved that the presence of linarin (L) is a strong chemophenetic marker, and automatically confirms the presence of *T. cordata* flowers in the analyzed sample. The presence of linarin (L) was observed in samples TS1, TS6–TS9, TS11–TS13, TS15, TS16, TS18, TS19, TS21–TS23, and TS26–TS30 (Supplementary materials Figs. S3 and S5). Therefore, these herbal products were classified as containing *T. cordata* flowers (Table 1). Based on the performed analyses it was clear that only samples TS6, TS15, TS25-28, and TS30 contained pure *T. cordata* flowers as no spots characteristic for other species were observed (Figs. S2–S5). Samples TS1, TS7, TS9, TS11–13, TS16, TS18, TS19, TS22, TS23, and TS29 consisted of herbal mixtures with the predominant content of *T. cordata* (except for TS16 and TS23 were *T. platyphyllos* or *T. europaea* flowers were major constituents). It was confirmed based on the observation of spots b, c, f, L, j, p, and q (Figs. S2–S5). Additionally, sample TS6 and 8 showed spot d (Fig. S2) proving that TS6 and TS8 are mixtures of *T. americana* + *T. cordata* with predominant *T. americana* in TS8 and *T. cordata* in TS6. Samples TS2, TS2, TS5, TS14, showed HPTLC profiles characteristic for *T. americana* (Figs. S3 and 1) with the presence of spots a, b, d, e, f, and absence of compounds L, j, m, n, p, and q. Thus, it was concluded that these samples exclusively contain flowers of American linden. Three samples (TS10, TS20, and TS21) were characterized as either *T. platyphyllos* or *T. europaea* without the addition of other species (Figs. S3 and S5). Finally, samples TS17 and TS24 were identified as *T. tomentosa* due to the presence of spots a, b, f, k and n, and a lack of other characteristic markers. In one sample (TS4), which was a complex herbal mixture product (Bronchisan), the identification of linden flowers used for production was not possible.

3.5. Chemometric analysis of obtained results – Development of SVM model

Due to the relatively low ratio of objects to features of the original data set (33 vs 13), the number of the latter had to be reduced in order to avoid the risk of overfitting the model. As shown in Fig. 4, some of the variables did not contain information useful for discriminating between the categories, while others, e.g. 0.82 and 0.51, are characteristic for a particular class of samples. The features selected based on the analysis of variance were, in order of decreasing relevance: 0.41, 0.51, 0.38, 0.25, 0.33, 0.82, and 0.74. Hierarchical clustering of the curtailed data set, shown in Fig. 4, revealed that the objects form three clusters at a height

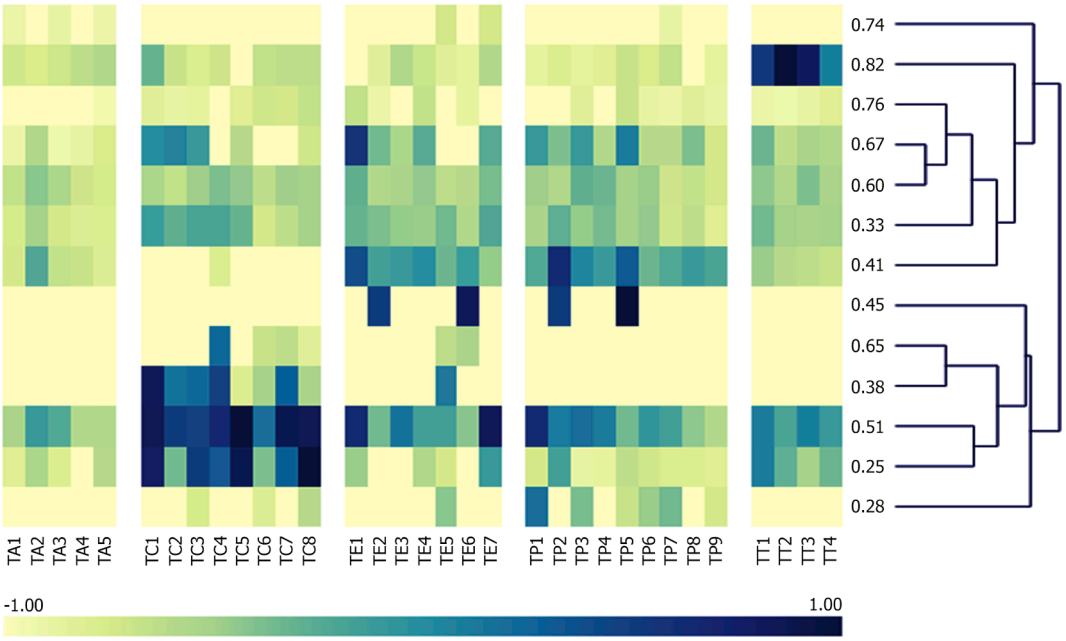


Fig. 4. Heat map depicting the normalized values grouped by class (TA – *T. americana*, TC – *T. cordata*, TE – *T* × *europaea*, TP – *T. platyphyllos*, and TT – *T. tomentosa*), with clustered variables.

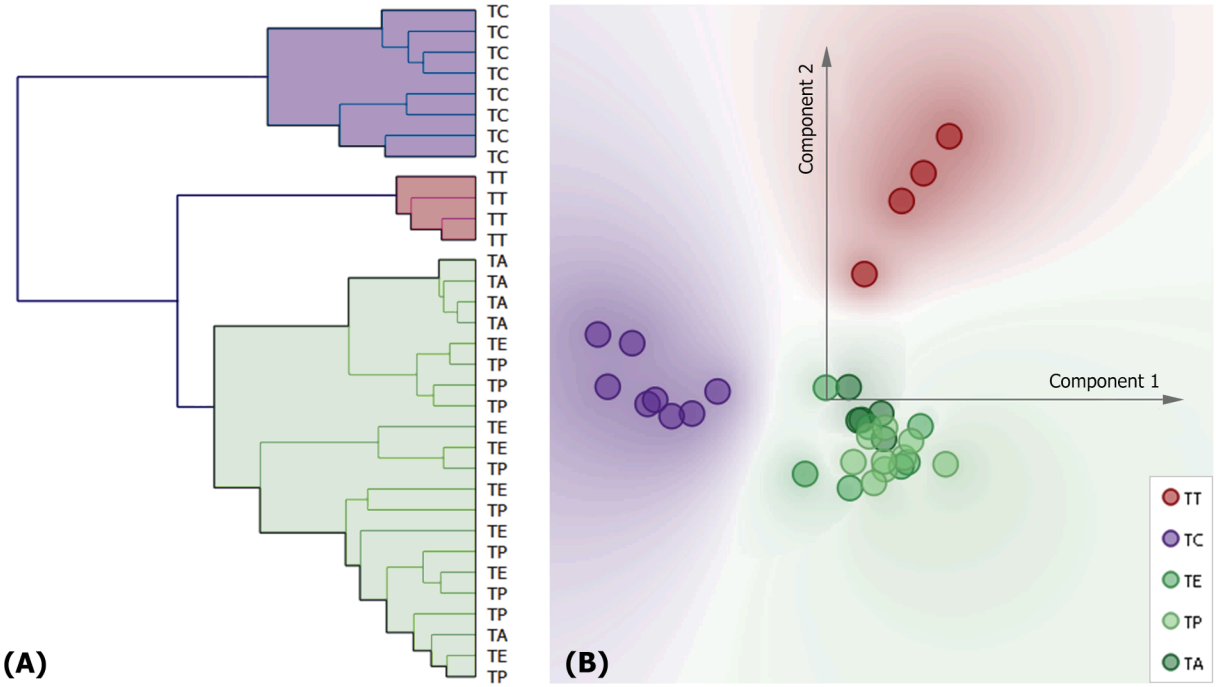


Fig. 5. A dendrogram obtained using hierarchical clustering with Ward linkage based on Euclidean distances, with clusters selected based on a 66% height ratio cut-off (A), and biplot of two first LDA components (B) (TA – *T. americana*, TC – *T. cordata*, TE – *T* × *europaea*, TP – *T. platyphyllos*, and TT – *T. tomentosa*).

ratio of 66%: two distinct clusters containing the TC and TT samples, respectively, and a cluster containing the TA, TE, and TP samples. This indicated that the latter three might be difficult to distinguish based on the input data. This observation is supported by the results of the LDA (Fig. 5). Due to the relatively small size of the TT group, the dimensionality of the dataset had to be further reduced from the initial seven to four variables, since in LDA the number of features should not, in general, exceed the number of objects in the least-numerous class (Tabachnick & Fidell, 2006). This was achieved using PCA, with the sum of the four first principal components covering 90% of the variance (44%, 18%, 15%, and 13% for PC1, PC2, PC3, and PC4, respectively). The 4 PCs were then used as inputs for LDA.

The seven features selected using the ReliefF algorithm were then used as predictors for an SVM-based machine learning model. The model's overall classification accuracy when considering all five separate classes was limited: 71% based on separate training and testing sets, and 76% based on cross-validation, with the former being an overall more robust approach to validation when dealing with relatively small datasets. This was to be expected based on the overlap of the TA, TE, and TP samples (Fig. 5). The model trained to predict all five distinct classes

(TT, TC, TA, TE, and TP) was subsequently used to classify the TS samples (see Table 1). It should be noted that, due to the differences between the TS and the validation data sets, and to the size of the training set, the SVM-based classification is not optimal. However, if the described approach were to be used in parallel with established methods, the classification accuracy of the machine learning model would gradually improve as the database established through routine tests would grow, eventually facilitating rapid screening and identification of suspect samples, which could be subjected to rigorous analysis.

The chemometric approach confirmed several observations pointed out after the visual analysis of HPTLC results. The HCA heat map conformed that TC contains significant amounts of spots at R_f ca. 0.25 and 0.65. It was also shown that the presence of linarin (L) at ca. 0.38 was confirmed only in TC samples and one sample of TE (TE5). This indicates that linarin should be considered as a strong chemophenetic marker for TC (Zidorn, 2019). Linarin was also identified as an important marker in our previous study using the UHPLC-DAD-MS approach (Ziaja et al., 2020). The present research confirms that this relatively rare flavonoid glycoside may be used for the authentication of extracts obtained from small-leaved lime. It was also shown that the presence of high amounts of *trans*-tiliroside is characteristic for samples of *T. tomentosa* (intensive spot at R_f ca. 0.82 was observed, Fig. 2). This observation was also confirmed in UHPLC analysis (Ziaja et al., 2020). The presence of intensive spots at ca. 0.41 and 0.45 was crucial for discrimination of TP and TE in the present study. Using the analysis of standards, a compound at ca. 0.41 was identified to be isoquercitrin (Fig. 2).

Based on SVM analysis each sample was classified as one of five investigated species (Table 1). The SVM algorithm was not capable of identifying herbal mixtures. It was assumed that the major species present in the analysis can only be tentatively identified. The analysis showed that all TS samples were categorized into three species (TC, TP, or TA). None of the extracts were identified as containing mainly TE or TT. Samples TS1–3, TS6–10, TS13, TS15 TS17, and TS19–30 were assigned as *T. cordata*. Samples TS5, TS14, TS16, and TS18 were identified as *T. platyphyllos*. Herbal teas TS4, TS11, and TS12 were authenticated as *T. americana* (Table 1).

3.6. Discrepancies and similarities between classification based on visual evaluation and SVM approach

Several inaccuracies between identified results obtained from visual analysis and SVM were revealed. Additionally, in some cases the manufacturer declaration of which species were used for the production of the analyzed herbal product was provided (Table 1). However, in several cases results from both methods were coherent. The analysis of samples TS1 and TS13 confirmed in both cases that it consists of mainly TC flowers. This was also declared by the supplier. Although both methods allowed for the detection of *T. americana* in sample TS4, visual evaluation proved that it is a minor constituent and TS4 mainly contains *T. cordata* (Table 1). Samples TS25-30 were identified in both models as mainly *T. cordata*.

In the case of several samples, significant inconsistencies between methods were observed. Samples TS2 and 3 were identified as TA in the visual evaluation and as TC using SVM approach. Additionally, the manufacturer declared that these products consist of TP flowers. A similar situation was encountered for samples TS14 and TS24.

Sample TS5 was characterized as TA based on observation of HPTLC results and TP by SVM. The latter was with the agreement with the producer declaration (Table 1).

The major problem with comparing the identification results obtained by both approaches can be due to the fact that the SVM analysis was not trained to identify herbal mixtures. Also, the number of samples for the algorithm learning was too small for obtaining reliable results. However, it was shown that the application of machine learning if further developed can be used for the identification of mixtures of linden flowers.

4. Conclusion

The current study showed that extracts from linden flowers obtained from most common species occurring in Europe can be easily distinguished using HPTLC fingerprints analysis. A fast, sensitive, and cheap chromatographic method was developed and used for the identification of species contained in several herbal products purchased in several European countries. The proposed methodology can be easily introduced for the quality control or authentication of linden flowers which is required by the Pharmacopoeia.

CRediT authorship contribution statement

Natalia Melnyk: Investigation, Data curation, Visualization. **Karolina A. Pawłowska:** Investigation, Data curation. **Maria Ziaja:** Investigation, Writing - review & editing, Visualization. **Wojciech Wojnowski:** Investigation, Data curation, Writing - review & editing. **Oleh Koshovyi:** Writing - review & editing. **Sebastian Granica:** Conceptualization, Methodology, Resources, Writing - review & editing, Project administration. **Agnieszka Bazylko:** Conceptualization, Methodology, Resources, Writing - review & editing, Data curation.

Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgements

The project was financially supported by the Polish National Science Centre research grant SONATA No. 2015/17/D/NZ7/02641. The authors wish to thank Roman Rudecki, Monika Marciniak, and Sebastian Doktór (all Warsaw) for their technical assistance. This project was carried out with the use of CePT infrastructure financed by the European Regional Development Fund within the Operational Programme "Innovative economy" for 2007–2013.

Appendix A. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.foodchem.2020.128929.

References

2016. European Pharmacopoeia, 9th ed. Strasbourg: Council of Europe.
Chang, C.-C., & Lin, C.-J. (2007). LIBSVM: A library for support vector machines. *ACM transactions on intelligent systems and technology 2*.
Demšar, J., Curk, T., Erjavec, A., Gorup, C., Hočevar, T., Milutinovič, M., … Zupan, B. (2013). Orange: Data mining toolbox in python. *Journal of Machine Learning Research, 14*, 2349–2353.
EMA/HMPC/337066/2011. (2012). Community herbal monograph on Tilia cordata Miller, Tilia platyphyllos Scop., Tilia x vulgaris Heyne or their mixtures, flos.
EMA/HMPC/346780/2011. (2011). Assessment report on Tilia tomentosa Moench, flos.
Evans, W. C., Evans, D., & Trease, G. E. (2009). *Trease and evans pharmacognosy*. Saunders/Elsevier.
Heinrich, M., Barnes, J., Prieto Garcia, J. M., Gibbons, S., Williamson, E. M., Kinghorn, A. D., & Blumenthal, M. (2018). *Fundamentals of pharmacognosy and phytotherapy*.
Karioti, A., Chiarabini, L., Alachkar, A., Fawaz Chehna, M., Vincieri, F. F., & Bilia, A. R. (2014). HPLC-DAD and HPLC-ESI-MS analyses of *Tiliae flos* and its preparations. *J. Pharma. Biomed. Anal., 100*, 205–214.
Pedregosa, F., Varoquaux, G., Gramfort, A., Michel, V., Thirion, B., Grisel, O., … Duchesnay, É. (2011). Scikit-learn: Machine learning in Python. *Journal of Machine Learning Research, 12*, 2825–2830.
Pigott, D. (2012). *Lime-trees and basswoods: A biological monograph of the genus Tilia*. Cambridge: Cambridge University Press.
Reich, E., & Schibli, A. (2007). *Dedication* (2007 ed.). Stuttgart: Georg Thieme Verlag.
Rutkowski, L. (2014). *Klucz do oznaczania roślin naczyniowych Polski niżowej*. Warszawa: Wydawnictwo Naukowe PWN.
Snyder, L. R. (1971). Solvent selectivity in adsorption chromatography on alumina: Non-donor solvents and solutes. *Journal of Chromatography A, 63*, 15–44.

Snyder, L. R. (1978). Classification off the solvent properties of common liquids. *Journal of Chromatographic Science, 16*, 223–234.
Tabachnick, B. G., & Fidell, L. S. (2006). *Using multivariate statistics* (5th ed.). Allyn & Bacon Inc.
Todorov, A. (2016). *An overview of the RELIEF algorithm and advancements, statistical approaches to gene X environment interactions for complex phenotypes*. The MIT Press.
Ziaja, M., Pawłowska, K. A., Józefczyk, K., Pruś, A., Stefańska, J., & Granica, S. (2020). UHPLC-DAD-MS/MS analysis of extracts from linden flowers (*Tiliae flos*): Differences in the chemical composition between five *Tilia* species growing in Europe. *Industrial Crops and Products, 154*, Article 112691 [in press].
Zidorn, C. (2019). Plant chemophenetics – A new term for plant chemosystematics/plant chemotaxonomy in the macro-molecular era. *Phytochemistry, 163*, 147–148.