# Biosynthesis of spathulenol and camphor stand as a competitive route to artemisinin production as revealed by a new chemometric convergence approach based on nine locations' field-grown *Artemesia annua* L.

Smain CHEMAT,[1] * Sonia BOUDJELAL,[1] Issam Malki,[2] Alexei LAPKIN [3] *

[1] *Extraction & Separation Techniques Team, Centre de Recherches Scientifique et Technique en Analyses Physico-Chimiques (C.R.A.P.C), BP 384, Zone Industrielle de Bousmail RP 42004 Tipaza, Algeria*

[2] Department of Economics and Quantitative Methods, University of Westminster, London W1B 2HW, United Kingdom

[3] Department of Chemical Engineering and Biotechnology, University of Cambridge, Cambridge, CB2 3RA, United Kingdom

* Corresponding authors: Smain CHEMAT, email: chemats@yahoo.fr; fax: +21324325775
Alexei LAPKIN, email: aal35@cam.ac.uk; fax: +441223334796

**Abstract**

Since isopentenyl diphosphate (IPP) and its isomer dimethylallyl diphosphate (DMAPP) are the universal precursors of both essential oil components, and the antimalarial agent artemisinin and its derivatives in *Artemesia annua* L., this paper aims to correlate the spotted differences in their concentrations by screening *Artemesia annua* L. field-grown in nine locations around the world that may reveal the role of any these compounds as precursors or competitors in the biosynthetic pathway of the sesquiterpene lactone : artemisinin.

Principal component analysis (PCA) revealed that artemisinin is positively correlated to β-pinene, 1,8-cineole, sabinene hydrate, borneol and 1-octen-3-ol; but negatively to artemisinic acid and β-caryophyllene oxide. Hierarchical cluster analysis (HCA) classified locations into two distinct groups in which artemisinin concentration stood as the main driving factor to build similarities between the locations.

In parallel, an improved convergence approach based on idiosyncratic similarities able to capture heterogeneity across individuals is proposed, which was able to classify compounds into four distinct clusters. Artemisinin appeared to be cross-linked to p-cymene, cis-carvyle acetate, 4-terpinene-1-ol, β-caryophyllene, β-farnesene, β-selinene, α-selinene, β-caryophyllene oxide and α-costol. It is interesting to see how camphor and spathulenol behaved as a distinct cluster group, which suggests that biosynthesis of these two compounds follows a different but a competitive pathway ; thus limiting their production could be a key to control and enhance the production of artemisinin.

**Keywords :** artemisinin, Malaria, Artemisia annua, essential oil, cluster, biosynthetic pathway

**Highlights**

- New idiosyncratic clustering approach is tested for *Artemesia annua* L. samples
- artemisinic acid and 32 identified constituents of essential oil are correlated to artemisinin
- camphor and spathulenol behaved as a distinct cluster and competitive to artemsinin biosynthesis

## 1. Introduction

Several studies have revealed the potency of *Artemisia* species in various traditional and folk medicines as a treatment for fever and malaria (Ortet et al., 2008 ; Tu, 2011). Particularly, artemisinin isolated from the traditional Chinese herb *Artemisia annua* serves as a precursor to today's most effective antimalarial drugs against strains of *Plasmodium falciparum* parasites (Meshnick et al., 1996). Moreover, recent reports suggest the use of Artemisia annua dried leaf tablets to treat resistant malaria in which the synergic role of other components with artemisinin is claimed to tackle plasmodium resistance (Daddy et al., 2017). However, production and accumulation of this endoperoxidized sesquiterpene lactone in the plant occurs at relatively low levels (0.01-1.4%). Therefore, understanding the biosynthetic pathway of the drug may help researchers to either tune its production in plant cells or mimic its biosythetic pathway into industrial synthetic routes.

Despite the great diversity of terpenoids in plant kingdom, they derive from a common biosynthetic precursor isopentenyl diphosphate (IPP) (Bouwmeester et al., 1999). Once IPP is formed and available in the cytosol, the production of farnesyl diphosphate (FDP) occurs via FDP synthase (FPS). The next step involves the formation of amorpha-4,11-diene via the sesquiterpene cyclase, amorpha-4,11-diene synthase (ADS), which represents the first specific precursor of artemisinin (Kim et al. 2006; Kim et al., 2008; Wallaart et al., 2001). Dihydroartemisinic acid (DHAA) is produced through three successive reactions catalyzed *in vitro* by cytochrome P450, CYP71AV1 (Teoh et al., 2006). The first reaction includes hydroxylation of amorpha-4,11-diene to form artemisinic alcohol, followed by an oxidation of the alcohol into artemisinic aldehyde. The latter is then reduced to dihydroartemisinic aldehyde and a final oxidation converts the aldehyde into the acid form (Wallaart et al., 1999). Finally, conversion of dihydroartemisinic acid (DHAA) into artemisinin is believed to be a non-enzymatic photo-oxidative reaction through the intermediate dihydroartemisinic acid peroxide (Wallaart et al., 2000). In parallel, the pathway branches also at artemisinic aldehyde to give artemisinic acid (AA) by the action of CYP71AV1 and/or ALDH1, and arteannuin B (AB) (Brown and Sy, 2007).

Biosynthetic routes for artemisinin and its derivatives have been well investigated but their relationships to other secondary metabolites such as essential oil components are scarce or unknown. Since isopentenyl diphosphate (IPP) and its isomer dimethylallyl diphosphate (DMAPP) are universal biosynthetic precursors of terpene essential oil constituents via the intervention of terpene synthases (TPS) (Rehman et al. 2016), their intervention may lead to favour competing routes; thus the accumulation or the limitation of certain class of compounds. Hence, chemometrics stand as an important tool to clarify which biosynthetic pathway is favoured upon an accumulation/production of specific metabolites, in this case essential oils.

Principal component analysis (PCA) and hierarchical clustering analysis (HCA) have become important tools to catch underlying similarities and classify groups using a simplified representation of correlations between samples/variables. Çam et al. (2009) succeeded to classify the chemotype of pomegranate cultivar based on the antioxidant activities of their juices, while Hossaina et al. (2011) used HCA to point out which polyphenolic compounds are most responsible for the *in vitro* antioxidant activity. Notable are studies of Suberu et al. (2016) using multivariate data analysis by discriminant function analysis (DFA) of the 13 varieties of *Artemisia annua* grown in Madagascar that helped them to identify a strong positive association between artemisinin and dihydroartemisinic acid. They suggested the occurrence of two different chemotypes in *A. annua*: one of high artemisinin and low artemisinic acid content and another one pertaining to those with low artemisinin and high artemisinic acid levels. However, PCA's ability to separate variations produced by each factor is under scrutiny due to the large spread across different number of components, while HCA does not provide explanation on how clusters are constructed and takes only elements with the smallest distance between each other (i.e. the most similar elements) to construct the cluster.

Phillips and Sul (2007) developed a novel convergence test approach able to capture heterogeneity across individuals using a varying factor-loading coefficient, which allows an endogenous identification of convergences and clustering variables into sub-groups. It displays the relative transition path for individuals, thus enabling to measure and capture the divergent behaviour of the individuals from the common stochastic trend that is varying across samples. This new clustering approach can be useful to detect relationships between secondary metabolites such as terpenoids and help us understand how competing biosynthetic pathways take place.

This study aims to explore relationships of artemisinin and artemisinic acid with essential oil compounds based on extraction yields data collected from a set of nine different locations of an identical *Artemesia annua* genetic origin. An improved clustering convergence approach developed by Phillips and Sul (2007) is applied to reveal compounds that are connected to higher yields of artemisinin or eventually those favouring other competing routes so their accumulation could be associated with low yields of artemisinin. Using regression analytics, a model is constructed to capture the impact of changes of competing secondary metabolites on artemisinin yields.

## 2. Material and Methods

### 2.1. Plant material

Samples of *Artemesia annua* L. leaves are sourced from REAP East Africa (Kenya), GSK (Tasmania, Australia), SensaPharm Ltd (United Kingdom) and Mundo Sano Foundation (Argentina). The samples represent commercial varieties for which companies hold voucher specimens which are collected at optimal artemisinin concentration in the herbs (just before flowering). The leaves are received by courier as crushed leaves in sealed high density carton

containers. Samples are placed in storage room under 4 ºC until utilization. In total, 9 samples are studied and named as follows: Argentina Conflitusato (A), Tasmania (B), Argentina Puerto valle (C), Argentina Carpeda 7 (D), Argentina Carpeda 6 (E), United Kingdom (F), Argentina Garruchos (G), Argentina Carpeda 8 (H) and East Africa (I).

It is important to make it clear that this study does not intend to relate composition with changes in environmental or climate factors, which are evident in our case as these samples are sourced from global locations but our objective is to relate changes in artemisinin concentration with their terpene essential oils.

**2.2. Artemisinin quantification**

In order to estimate the total content of artemisinin and artemisinic acid in a given sample, extraction at 40 ºC of 20 g samples of *Artemesia annua L.* is performed with 200 mL of ethyl acetate and fresh portions are added twice every 3 hours. The three extracts are mixed and evaporated to dryness under vacuum and re-dissolved in 10 mL of acetonitrile. The obtained solutions are filtered through a 0.2 μm syringe filter before injection into HPLC. In this case, analysis of artemisinin and artemisinic acid is realized using a Shimadzu Prominence HPLC equipped with a UV-vis diode array detector (SPD-M20A, DAD) coupled to an evaporative light scattering detector (ELSD, LTII, 350 kPa $N_2$, nebulizer at 40 ºC). Column Betasil C18 column, set at a temperature of 40 ºC, is used with an acetonitrile:water (65:35 %v/v) mobile phase at an isocratic mode 0.8 mL $min^{-1}$ flow rate as described elsewhere (Lapkin et al. 2009). Calibration curves are constructed from injection into HPLC/ELSD of dilutions varying from 0.25 to 5 mg $mL^{-1}$ of standards of artemisinin provided by Neem Biotech Ltd. (Newport, Wales, UK), and artemisinic acid kindly provided by Walter Reed Army Institute of Research (Washington, USA).

**2.3. Essential oil composition**

A dry sample of 100 g of *Artemesia annua* L. is suspended in 1 L of water and boiled for a period of six hours in a modified Clevenger-type system that allows recirculation of the condensed water and easy collection of the essential oil fraction at the end. This fraction is collected in vials and submitted for composition analysis. Extractions are performed twice and the mean values are reported.

Essential oil extracts are analysed by gas chromatography coupled to mass spectrometry (GC-MS) (Hewlett-Packard 6890 GC coupled to a 5973A MS) using two fused-silica-capillary columns. The first one is a non-polar column HP5MS™ (30 m x 0.25 $mm^2$, 0.25 μm film thickness) and the second is Stabilwax™ polar column consisting of Carbowax™-PEG (60 m x 0.2 $mm^2$, 0.25 μm film thickness). GC-MS spectra are obtained using the following conditions: carrier gas He; flow rate 0.3 mL $min^{-1}$; split-less mode; injection volume 0.2 μL; injection temperature 250 °C; the oven

temperature programme is 60 °C for 8 min increased at 2 °C $min^{-1}$ to 250 °C then held at 250 °C for 15 min; the ionisation mode used is electronic impact at 70 eV.

The identification of essential oil components is achieved by comparison of GC Kovats retention indices (R.I.) of compounds, determined with reference to homologous series of $C_5$–$C_{28}$ *n*-alkanes, with those of authentic standards available in the authors' laboratory. The identification process is confirmed when possible by comparison of mass spectral fragmentation patterns with those stored in the MS database (National Institute of Standards and Technology and Wiley libraries) and with mass spectra data of literature (Bagchi et al., 2003; Tzenkova et al., 2010). Component relative molar concentrations are obtained directly from GC peak areas.

### 2.4. Statistical Data Analysis

Results on the metabolic profiling of different samples of *Artemesia annua* L. leaves are organized in a dataset where plant origin (individuals) is placed in lines while components (variables) are placed in columns in order to obtain pattern recognition analysis via descriptive statistics using *XLSTAT* software. A standardized principal component analysis (PCA) based on covariance (Pearson) is conducted in order to assess correlations between each component where variables are centred then reduced. In this case, a two dimensional (2D) factorial plan capturing the position of each compound relative to each other is created by depicting principal components' values relative to each location (Kusa et al., 2009). In these tests, the significance level at which we estimated the critical values, the differences were 5% (i.e. $P \leq 0.05$).

In an attempt to arrange plant origin, a hierarchical clustering analysis (HCA) based on Ward aggregation method is realized using *XLSAT* software in order to classify locations into groups according to their high correlations. Dissimilarity is measured by Euclidean distance for nine locations according to metabolic profile of each *A. annua* sample. The results are illustrated as a dendrogram.

### 2.5. Improved clustering approach

Phillips and Sul (2007) PS hereafter, propose an idiosyncratic element that is allowed to evolve over a variable factor and captures heterogeneity across individuals using a varying factor-loading coefficient by testing convergence, then identifying the endogenous convergence of clusters. According to them, the nonlinear transition factor model is defined as shown in equation 1:

$X = \delta$ Eq. (1)

where is the dependent variable observed across individuals (compounds in our case) that changes according to variable elements (locations in our case).

In other words, the coefficient measures the share of the common factor for each individual in the panel data experiments.

The analysis of convergence is based on the application of a relative transition coefficient, , which measures the loading coefficient in relation to the panel average for a variable *j*. The parameter is approximated by equation 2:

Eq. (2)

Using this parameter along with the loading coefficient $\delta_{it}$, the convergence can be assessed. Namely, if , then 1, thus the cross-sectional variance of converges to zero, and we have:

Eq. (3)

The property in equation 3 is essential in testing the null of convergence, then clustering individuals into the convergence clubs. The coefficient $h_{ij}$ displays the relative transition path for individuals in our panel data, then measures and captures the divergent behaviour of individuals from the common stochastic trend that is varying according to plant origin $\mu_j$.

The procedure is implemented in two stages. In the first stage, the null of overall convergence is tested, $H_0: \delta_i = \delta$ and $\alpha \geq 0$, against the alternative of no convergence, $H_0: \delta_i \neq \delta$ for all $i$, or 0. In order to test this hypothesis a *logarithm* regression is estimated, which is based on the cross sectional variance ratio as proposed by PS . The regression is defined as follows:

Eq. (4)

where

Eq. (5)

Phillips and Sul (2007) recommend to use $r$=0.3. Regression calculations are realised using IBM SPSS Statistics V.20 software. Once the regression is run, the null is accepted if the autocorrelation heteroskedasticity robust one tail $t_{\hat{b}}$ statistic is above the critical value, $c$ (e.g. at 5% level of significance, accept the null if $t_{\hat{b}} \geq -1.65$). In the second stage, another test relative to the presence of club convergence is realized.

**3. Results and Discussion**

A significant variation in terms of the relative amounts of artemisinin, artemisinic acid and major essential oil compounds is observed in the biomasses grown at different geographic regions, as shown in Table 1. Thirty-two (32) essential oil components are identified, with camphor as the

major compound ranging from 13.05 to 55.5% (Molar). Other major components of the essential oil include: spathulenol (2.3-8.05%), α-costol (2-5.8%), β-caryophylene oxide (1.45-5.35%), borneol (0.08-4.65%) and β-farnesene (0.31-3.2%). It is important to note that certain locations contain specific compounds such as cedrol, α-cedrene and α-bergamotene for United Kingdom (F), while α-curcumene is found only in East Africa (I), United Kingdom (F) and Tasmania (B). The richest location in terms of camphor yield belongs to Conflitusato Argentina (A) with 55.5%. Surprisingly, artemisia ketone could not be found in the examined essential oils despite being largely reported in the literature in a range of up to 26% (Radulović et *al.*, 2013).

Wang et al. (2009) spotted this composition difference in two studied essential oils where camphor, methyl artemisinic acid and lanceol were absent in one genotype, which had impacted negatively on the content in artemisinin but positively on the content of arteannuin B and dihydroartemisinic acid. These differences in metabolic profiles could be attributed to cultivation, environmental factors and harvesting period. Specifically, earlier studies have revealed significant variations of some flavonoids and pigments in the samples collected from different locations (Lapkin et *al.*, 2014). If pure artemisinin recovery is sought, tuning downstream processing protocols for raw materials sourced from different places becomes inevitable.

Certain locations are found to be rich in the oxygenated compounds like Conflitusato Argentina (A) (71.44%) followed by Carpeda 6 Agentina (E) (68.88%) then Puerto Valle Argentina (C) (68.79%). In parallel, Carpeda 7 Argentina (D) exhibited the highest artemisinin and artemisinic acid contents with 14.08 and 2.1 mg $g^{-1}$ of dry leaves respectively. Artemisinin yields are reported to range from 4.4 mg $g^{-1}$ (Mannan et *al.*, 2010) to up to 11.5 mg $g^{-1}$ (Ferreira et *al.*, 2013); these values are comparable to the studied varieties having yields from 4.09 to 14.08 mg $g^{-1}$.

**3.1. Data analysis**

Pattern classification results indicate that artemisinin and camphor variables represent the main source of dispersion because of their high mean values compared to other variables, and hold higher standard deviations between locations. Following this step, correlation data produced by standardized Pearson principal component analysis (PCA) shows that artemisinin holds high correlation factors, positive with artemisinic acid, β-pinene, 1,8-cineole, sabinene hydrate, borneol and 1-octen-3-ol, but negative with β-caryophyllene oxide. In parallel, very weak positive correlations were observed towards β-caryophyllene and chrysanthenyle acetate. For artemisinic acid, the data indicates that it is highly positively related to borneol. However, only 63% of the initial information could be explained by these correlations in the first factorial plans (F1 & F2). Therefore, a non-standardized PCA is conducted. The latter succeeded to explain more than 90% in a single factorial axis (F1) and reached more than 97% of correlation's information in the first factorial plan (F1-F2), see Figure 1. In this case, F1 axis appears to represent intimately the

camphor variable, while F2 axis seems to represent artemisinin well. The results show a congregation into three clusters where artemisinin appeared strongly linked to artemisinic acid, borneol, 1,8-cineole, β-pinene and σ-terpinene. Sesquiterpenes took the negative side of F1 axis but are divided in two clusters with a mix of oxygenated and hydrocarbon compounds in both groups. A high correlation of 78% exists between artemisinin and artemisinic acid, which can be translated by simple regression to include other secondary metabolites with whom artemisinin hold high correlation ($R^2$=0.965) into the following equation 6 :

Eq. (6)

Equation 6 means that an increase in artemisinic acid would not drive a substantial decrease in artemisinin yield. In other words, they seem to belong to the same but non-competitive biosynthetic pathway where artemisinic acid may be considered as a by-product of the artemisinin synthesis. Our results are in line with Wallaart et *al.* (2010) findings who noted a negative correlation between artemisinic acid and artemisinin. Suberu et *al.* (2016) also indicated that a higher accumulation of artemisinin can be associated with the lower concentrations of arteannuin B and artemisinic acid and vice versa. Nonetheless, these relationships remain strictly statistical unlike Brown and Sy's experimental work using *in vivo* feeding trials with artemisinic acid labelled with both $^{13}C$ and $^{2}H$ at the 15-position that indicated that the produced sesquiterpene metabolites retained their unsaturation at the 11,13-position, so excluding the conversion of artemisinin acid in *A. annua* into any 11,13-dihydro metabolite, including artemisinin (Brown and Sy, 2007).

According to Teoh et *al.* (2006), artemisinin and artemisinic acid are produced from the same precursor artemisinic aldehyde. This suggests that a reduction of artemisinic aldehyde into dihydroartemisinic aldehyde is not complete, which leaves a window for the production of artemisinic acid to take place (Vail, 2008). Furthermore, artemisinin is positively cross-linked with borneol and 1,8-cienole. Hence, favouring their production may boost the production of artemisinin. In parallel, artemisinin production seems to be negatively influenced by the occurrence of β-pinene and 1-octen-3-ol. This implies that these compounds may belong to a distinct biosynthetic route. This is partly true as the production of 1-octen-3-ol is non-enzymatic and does not come from IPP nor DMAPP.

In another part, aggregation of the most similar variables (compounds) across locations helped the construction of hierarchical ascendant clustering (HCA) in order to detect proximities between individuals (locations) as illustrated by a dendrogram in Figure 2. This figure revealed the presence of different clustering levels. The first level indicate that F and G locations belongs to the same chemotype, probably due to the similar yields of artemisinin, though with a different composition and spatial origin (United Kingdom and Argentina), while the rest of locations are believed to

belong to a distinct chemotype. Furthermore, locations of D, H, B and C hold strong similarities, attributed mainly to high artemisinin content. These observations should be taken with caution as these are only based on statistical characteristics that take short Euclidean distances between compounds as an indication of high similarity, thus a strong relationship is expected.

### 3.2. PS clustering approach

Another attempt to classify compounds using a PS approach helped the construction of clusters, see Table 2. The estimated t-statistics of the overall convergence test indicates that the null hypothesis is rejected but it does not imply that there is no convergence between a restricted number of compounds. It may, in fact, imply that we should move to a relative convergence test, in which the outcomes suggest that compounds can be grouped into four distinct clusters. Figure 3 shows graphically how the clusters are distributed across locations. A clear gap is observed between the first cluster, comprising camphor and spathulenol, and the remaining clusters that seems to be amassed (clusters 2-4). A magnification of clusters 2-4 allows to see the discrepancy between the aggregated clusters. It is possible now to spot the significant shift of the cluster 2 relative to the clusters 3 and 4 at the location G. This can be attributed to the high content of camphor in G. Nonetheless, a relatively regular pattern of clusters is recorded across other locations.

Camphor and spathulenol belong to a distinct cluster, which suggests that they are produced through a different biosynthetic route. In fact, spathulenol structure comprises a cycloheptanyl skeleton like artemisinin, which suggests it stands on a competitive pathway to artemisinin synthesis.

We may thus hypothesise that control of spathulenol and camphor biosynthesis may lead to maximize artemisinin yield in the plants. In parallel, artemisinin seems to be associated with the production of major sesquiterpene hydrocarbons, such as α-costol, β-caryophyllene, β-farnesene and β-caryophyllene oxide. Interestingly, the last three compounds are reported to be produced along with artemisinin parent “amorpha-4,11-diene” under enzymatic catalysis from the same precursor farnesyl diphosphate (Weathers et al., 2011), which may explain their aggregation with artemisinin in the same cluster. Artemisinin acid seems to be linked mainly to a mix of oxygenated and hydrocarbon monoterpenes, but essentially to the sesquiterpene Germacren D which shares the same skeleton with artemisinic acid. However, they derive from competing biosynthetic routes as Germacren D is produced via the cyclization of farnesyl diphosphate by Germacren A synthase, unlike artemisinin acid that is produced via cyclization of farnesyl disphosphate by amorphadiene synthase (Nguyen et al., 2010). This observation supports the proficiency of the PS clustering approach to provide distinct clusters.

## 4. Conclusions

Considerable variations are observed between *Artemesia annua* L. samples sourced from nine locations in terms of artemisinin, artemisinic acid and essential oil components in which Argentinean varieties hold the highest artemisinin yields.

The PCA statistical analysis enabled us to classify this complex dataset, but it has failed to match some of the underlying relationships between compounds with the available literature on biosynthetic routes. Hierarchical clustering analysis (HCA) indicated the presence of three levels of relationships between the nine studied locations, mainly arising from their relative content in artemisinin. In parallel, the idiosyncratic approach developed by Philips & Sul succeeded, to a large extent, to emphasise the role of sesquiterpene hydrocarbons in the biosynthetic route of artemisinin, while advocating a new approach to maximize artemisinin content through control of other terpenic compounds spathulenol and camphor production in the plants. This new information may contribute to elucidate controlling mechanisms of artemisinin biosynthesis, a key to an improved artemisinin biotechnology in the transgenic *A. annua*. In addition, the new convergence tool may evolve as support tool to reveal biosynthetic routes in other plants.

**Acknowledgments**

This work was funded by the Engineering and Physical Sciences Research Council, project "Adaptive Processing of Natural Feedstocks", EP/F016182/1. The authors would like to thank Professor F. CHEMAT from Université d'Avignon et de Pays de Vaucluse (France) for GC/MS analyses. HPLC instrument was purchased with the help of Medicines for Malaria Ventures (MMV).

**References**

Bagchi, G.D., Haider, F., Dwivedi, P.D., Singh, A., Naqvi, A.A., 2003. Essential oil constituents of *Artemisia annua* during different growth periods at monsoon conditions of subtropical north Indian plains. *J. Ess. Oil Res.*, 15(4), 248-250. https://doi.org/10.1080/10412905.2003.9712131

Bouwmeester, H.J., Wallaart, T.E., Janssen, M.H., Van Loo, B., Jansen, B.J., Posthumus, M.A., Schmidt, C.O., De Kraker, J.W., Konig, W.A., Franssen, M.C., 1999. Amorpha-4,11-diene synthase catalyses the first probable step in artemisinin biosynthesis. *Phytochem*., 52, 843-854. https://doi.org/10.1016/S0031-9422(99)00206-X

Brown, G.D., Sy, L.K., 2007. In vivo transformations of artemisinic acid in *Artemisia annua* plants. *Tetrahedron*, 63, 9548-9566. https://doi.org/10.1016/j.tet.2007.06.062

Çam, M., Hişil, Y., Durmaz, G., 2009. Classification of eight pomegranate juices based on antioxidant capacity measured by four methods. Food Chem., 112(3), 721-726. https://doi.org/10.1016/j.foodchem.2008.06.009

Daddy, N.B., Kalisya, L.M., Bagire, P.G., Watt, R.L., Towler, M.J., Weathers, P.J., 2017. Artemisia annua dried leaf tablets treated malaria resistant to ACT and i.v. artesunate: Case reports. Phytomedicine, 32(15), 37-40.

https://doi.org/10.1016/j.phymed.2017.04.006.

Ferreira, J.F.S., Zheljazkov, V. D., Gonzalez, J. M., 2013. Artemisinin concentration and antioxidant capacity of *Artemisia annua* distillation by-product. *Ind. Crops Prod.*, 41, 294-298. https://doi.org/10.1016/j.indcrop.2012.05.005

Hossaina, M.B., Patrasb, A., Barry-Ryana, C., Martin-Dianaa, A.B., Bruntonc, N.P., 2011. Application of principal component and hierarchical cluster analysis to classify different spices based on in vitro antioxidant activity and individual polyphenolic antioxidant compounds. *J. Funct. Foods*, 3, 179-189.
https://doi.org/10.1016/j.jff.2011.03.010

Kim, S.H., Heo, K., Chang, Y.J., Park, S.H., Rhee, S.K., Kim, S.U., 2006. Cyclization mechanism of amorpha-4,11-diene synthase, a key enzyme in artemisinin biosynthesis. *J. Nat. Prod.*, 69, 758-762. https://doi.org/10.1021/np050356u

Kim, S.H., Chang, Y.J., Kim, S.U., 2008. Tissue specificity and developmental pattern of amorpha-4,11-diene synthase (ADS) rroved by ADS promoter-driven GUS expression in the heterologous plant *Arabidopsis thaliana*. *Planta Med.*, 74, 188-193.
https://doi.org/10.1055/s-2008-1034276

Kusari, S., Zühlke, S., Borsch, T., Spiteller, M., 2009. Positive correlations between hypericin and putative precursors detected in the quantitative secondary metabolite spectrum of *Hypericum*. *Phytochem.*, 70, 1222-1232.
https://doi.org/10.1016/j.phytochem.2009.07.022

Lapkin, A.A., Walker, A., Sullivan, N., Khambay, B., Mlambo, B., Chemat, S., 2009. Development of HPLC analytical protocols for quantification of artemisinin in biomass and extracts. *J. Pharm. Biomed. Anal.*, 49(4), 908-915.
https://doi.org/10.1016/j.jpba.2009.01.025.

Lapkin, A., Adou, E., Mlambo, B.N., Chemat, S., Suberu, J., Collis, A.E.C., Clark, A., Barker, G., 2014. Integrating medicinal plants extraction into a high-value biorefinery: an example of *Artemisia annua* L. *C. R. Chim.*, 17, 232-241.
https://doi.org/10.1016/j.crci.2013.10.023

Mannan, A., Ahmed, I., Arshad, W., Asim, M.F., Qureshi, R.A., Hussain, I., Mirza, B., 2010. Survey of artemisinin production by diverse *Artemisia* species in northern Pakistan. *Malar. J.*, 9, 310-318.
https://doi.org/10.1186/1475-2875-9-310

Meshnick, S.R., Taylor, T.E., Kamchonwongpaisan, S., 1996. Artemisinin and the antimalarial endoperoxides: from herbal remedy to targeted chemotherapy. *Microbiol. Rev.*, 60(2), 301-315.

Nguyen, D.T., Göpfert, J.C., Ikezawa, N., MacNevin, G., Kathiresan, M., Conrad, J., Spring, O., Ro, D-K., 2010. Biochemical Conservation and Evolution of Germacrene A Oxidase in Asteraceae. *The J. of Biol. Chem.*, 285(22)16588–16598
https://doi.org/10.1074/jbc.M110.111757

Ortet, R., Prado, S., Mouray, E., Thomas, O.P., 2008. Sesquiterpene lactones from the endemic Cape Verdean Artemisia gorgonum. *Phytochem.*, 69(17), 2961-2965. https://doi.org/10.1016/j.phytochem.2008.09.022

Phillips, P.C.B., Sul, D., 2007. Transition modelling and econometric convergence tests. *Econometrica.*, 75(6), 1771-1855. https://doi.org/10.1111/j.1468-0262.2007.00811.x

Radulović, N.S., Randjelović, P.J., Stojanović, N.M., Blagojević, P.D., Stojanović-Radić, Z.Z., Ilić, I.R., Djordjević, V.B., 2013. Toxic essential oils. Part II: chemical, toxicological, pharmacological and microbiological profiles of Artemisia annua L. volatiles. *Food Chem. Toxicol.*, 58, 37-49. https://doi.org/10.1016/j.fct.2013.04.016

Rehman, R., Hanif, M.A., Mushtaq, Z., Al-Sadi, A.M., 2016. Biosynthesis of essential oils in aromatic plants: A review. *Food Rev. Int.* 32(2),117-160 https://doi.org/10.1080/87559129.2015.1057841

Suberu, J., Gromski, P.S., Nordonb, A., Lapkin, A., 2016. Multivariate data analysis and metabolic profiling of artemisinin and related compounds in high yielding varieties of *Artemisia annua* field-grown in Madagascar. *J. of Pharmaceutical Biomed. Anal.*, 117, 522-531. https://doi.org/10.1016/j.jpba.2015.10.003

Teoh, K.H., Polichuk, D.R., Reed, D.W., Nowak, G., Covello, P.S., 2006. *Artemisia annua* L. (Asteraceae) trichome-specific cDNAs reveal CYP71AV1, a cytochrome P450 with a key role in the biosynthesis of the antimalarial sesquiterpene lactone artemisinin. *FEBS Lett.*, 580(5), 1411-1416. https://doi.org/10.1016/j.febslet.2006.01.065

Tu, Y., 2011. The discovery of artemisinin (qinghaosu) and gifts from Chinese medicine. *Nat. Med.*, 17, 1217-1220. https://doi.org/10.1038/nm.2471

Tzenkova, R., Kamenarska, Z., Draganov, A., Atanassov, A., 2010. Composition of *Artemisia annua* essential oil obtained from species growing wild in Bulgaria. *Biotechnol. Biotechnol.*, 24(2), 1833-1835. https://doi.org/10.2478/V10133-010-0030-6

Vail, D.R., 2008. Artemisinin Biosynthesis: Developmental and Sugar Regulation of mRNA Levels. Master thesis, *Worcester Polytechnic Institute*, MA.

Wallaart, T.E., van Uden, W., Lubberink, H.G., Woerdenbag, H.J., Pras, N., Quax, W.J., 1999. Isolation and identification of dihydroartemisinic acid from *Artemisia annua* and its possible role in the biosynthesis of artemisinin. *J. Nat. Prod.*, 62, 430-433. https://doi.org/10.1021/np980370p

Wallaart, T.E., Pras, N., Beekman, A.R.C., Quax, W.J., 2000. Seasonal variation of artemisinin and its biosynthetic precursors in plants of *Artemisia annua* of different geographical origin: proof for the existence of chemotypes. *Planta Med.*, 66, 57-62. https://doi.org/10.1055/s-2000-11115

Wallaart, T.E., Boumeester, H.J., Hill, J., Maijres, C.A., 2001. Amorpha-diene synthase cloning and functional expression of a key enzyme in the biosynthetic pathway of a novel antimalarial drug Artemisinin. *Planta.*, 212, 460-465.
https://doi.org/10.1007/s004250000428

Wang, H., Ma, C., Ma, L., Du, Z., Ye, H., Li, G., Liu, B., Xu, G., 2009. Secondary metabolic profiling and artemisinin biosynthesis of two genotypes of *Artemisia annua*. *Planta Med.*, 75, 1625-1633.
https://doi.org/10.1055/s-0029-1185814

Weathers, P.J., Arsenault, P.R., Covello, P.S., McMickle, A., Teoh, K.H., Reed, D.W., 2011. Artemisinin production in *Artemisia annua*: studies in planta and results of a novel delivery method for treating malaria & other neglected diseases. *Phytochem. Rev*., 10(2), 173-183.
https://doi.org/10.1007/s11101-010-9166-0

**Table 1.** Artemisinin, artemisinic acid and essential oils composition of *Artemesia annua* L. from nine locations

| | Composition (%) | RI[†] | Conflitusato (Argentina) **(A)** | Tasmania **(B)** | Puerto valle (Argentina) **(C)** | Carpeda 7 (Argentina) **(D)** | Carpeda 6 (Argentina) **(E)** | UK **(F)** | Garruchos (Argentina) **(G)** | Carpeda 8 (Argentina) **(H)** | East Africa **(I)** |
|---|---|---|---|---|---|---|---|---|---|---|---|
| | Artemisinin (mg/g) | - | 7.82 | 9.69 | 11.27 | 14.08 | 9.31 | 3.29 | 3.28 | 12.95 | 4.09 |
| | Artemisinic acid (mg/g) | - | 0.98 | 1.14 | 0.82 | 2.10 | 1.23 | 0.63 | 0.35 | 1.20 | 0.53 |
| | **Monoterpene hydrocarbons** | | | | | | | | | | |
| 1 | α-pinene | 936 | 0.41 | 0.17 | 0.18 | 0.31 | 0.08 | 0.01 | 0.15 | 0.47 | 0.01 |
| 2 | camphene | 941 | 1.10 | 1.32 | 1.22 | 2.05 | 0.28 | t[‡] | 0.61 | 3.00 | 0.05 |
| 3 | β-pinene | 974 | 0.10 | 0.05 | 0.05 | 0.15 | 0.03 | t[‡] | 0.01 | 0.13 | t[‡] |
| 4 | *p*-cymene | 1019 | 1.50 | 0.80 | 0.65 | 1.25 | 0.88 | 0.05 | 0.91 | 1.15 | 0.27 |
| 5 | Limonene | 1023 | 0.01 | 0.07 | 0.02 | 0.12 | 0.02 | 1.95 | 0.09 | 0.06 | t[‡] |
| 6 | σ-terpinene | - | 0.15 | 0.10 | 0.10 | 0.18 | 0.06 | 0.10 | 0.08 | 0.22 | 0.06 |
| | **Oxygenated Monoterpenes** | | | | | | | | | | |
| 7 | 1,8-cineole | 1025 | 1.50 | 1.15 | 2.05 | 2.45 | 0.65 | 0.23 | 0.72 | 3.00 | 0.37 |
| 8 | Sabinene hydrate <*cis*> | 1062 | 0.35 | 0.38 | 0.40 | 0.35 | 0.25 | 0.02 | 0.08 | 0.25 | 0.02 |
| 9 | Chrysantenone | - | 0.13 | 0.14 | 0.36 | 0.18 | 0.23 | 0.03 | 0.20 | 0.41 | 0.03 |
| 10 | Chrysanthenyl acetate | - | 0.31 | 0.22 | 0.20 | 0.29 | 0.32 | 0.62 | 0.27 | 0.58 | 0.03 |
| 11 | 1,8-dihydrocineole | – | 0.65 | 0.20 | 0.10 | 0.15 | 0.55 | 0.14 | 0.45 | 0.12 | 1.20 |
| 12 | Carvyl acetate <*cis*> | - | 1.10 | 0.18 | 0.10 | 0.15 | 1.05 | 0.05 | 1.15 | 0.12 | 0.58 |
| 13 | camphor | 1129 | 55.50 | 49.85 | 52.30 | 50.90 | 46.40 | 43.05 | 47.15 | 47.75 | 44.10 |
| 14 | pinocarvone | 1151 | 0.29 | 0.18 | 0.18 | 0.21 | 0.25 | 0.18 | 0.22 | 0.34 | 0.20 |
| 15 | Borneol | 1161 | 1.90 | 3.30 | 2.95 | 4.65 | 2.85 | 0.08 | 0.31 | 4.10 | 0.68 |
| 16 | Terpinen-4-ol | 1172 | 0.80 | 0.42 | 0.55 | 0.75 | 0.70 | 0.70 | 0.78 | 0.80 | 1.10 |
| 17 | α-terpineol | 1203 | 0.23 | 0.21 | 0.25 | 0.31 | 0.25 | 0.19 | 0.22 | 0.30 | 0.26 |
| 18 | Carveol <*cis*> | 1219 | 0.28 | 0.05 | 0.05 | 0.19 | 0.28 | 0.05 | 0.10 | 0.25 | 0.01 |

| | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|
| | **Sesquiterpene hydrocarbons** | | | | | | | | | | |
| 19 | α-copaene | 1375 | 0.50 | 0.53 | 0.55 | 0.52 | 0.25 | 0.60 | 0.69 | 0.52 | 0.45 |
| 20 | α-cedrene | | - | | - | - | | 0.45 | | - | - |
| 21 | β-caryophyllene | 1421 | 1.80 | 1.32 | 2.00 | 2.30 | 1.60 | 2.20 | 1.45 | 2.30 | 2.70 |
| 22 | β-farnesene | 1456 | 0.70 | 0.31 | 0.75 | 0.85 | 0.80 | 1.30 | 1.05 | 0.90 | 3.20 |
| 23 | α-curcumene | - | - | 2.50 | - | - | - | 0.20 | - | - | 3.15 |
| 24 | α-bergamotene | - | - | - | - | - | - | 2.50 | - | - | - |
| 25 | Germacren D | 1468 | 1.75 | 1.00 | 1.85 | 2.25 | 2.25 | 0.70 | 0.67 | 2.40 | 2.40 |
| 26 | β-selinene | 1475 | 0.85 | 0.45 | 1.10 | 0.95 | 0.77 | 2.55 | 1.10 | 1.05 | 1.30 |
| 27 | α-selinene | - | 0.50 | 0.45 | 0.55 | 0.65 | 0.78 | 0.55 | 0.85 | 0.65 | 0.85 |
| | **Oxygenated sesquiterpenes** | | | | | | | | | | |
| 28 | β-caryophyllene oxide | 1576 | 2.10 | 2.70 | 2.30 | 1.95 | 1.45 | 5.35 | 4.00 | 1.65 | 4.40 |
| 29 | Spathulenol | 1564 | 4.05 | 2.30 | 4.05 | 3.60 | 7.00 | 8.05 | 6.55 | 3.55 | 5.00 |
| 30 | α-costol | - | 2.00 | 3.05 | 2.95 | 2.75 | 3.05 | 5.80 | 2.55 | 2.45 | 4.50 |
| 31 | cedrol | - | - | - | - | - | - | 0.85 | - | - | - |
| | **Others** | | | | | | | | | | |
| 32 | 1-octen-3-ol | 973 | 0.25 | 0.28 | 0.20 | 0.30 | 0.20 | - | 0.10 | 0.55 | 0.06 |
| | **Total oxygenated compounds** | - | 71.44 | 64.61 | 68.79 | 68.88 | 65.28 | 65.39 | 64.95 | 65.67 | 62.48 |
| | **Total non-oxygenated compounds** | - | 9.37 | 9.07 | 9.02 | 11.58 | 7.80 | 13.16 | 7.35 | 12.85 | 14.44 |
| | **Total Unidentified compounds** | - | 18.96 | 26.04 | 21.99 | | 26.72 | | 27.60 | 20.93 | 23.02 |

†: retention indices calculated on DB1-column

‡: t: traces

520 **Table 2.** Clustering results of the convergence test run according to Phillips and Sul (2007).
521

| | Overall Test | Cluster 1 | Cluster 2 | Cluster 3 | Cluster 4 |
|---|---|---|---|---|---|
| Estimated *t*-statistics* | [-3.92<-1.65] | [***-1.48***>-1.65] | [***-1.57***>-1.65] | [***-1.53***>-1.65] | [***-1.50***>-1.65] |
| Composition | no convergence | Camphor<br>Spathulenol | Artemisinin<br>p-Cymene<br>cis-Carvyle acetate<br>4-Terpinene-1-ol<br>β-Caryophyllene<br>β-Farnesene<br>β-Selinene<br>α-Selinene<br>α-costol<br>β-Caryophyllene oxide | Artemisinic acid<br>Camphene<br>1,8-Cineole<br>Pinocarvone<br>α-Terpineol<br>1,8-Dihydrocineole<br>Borneol<br>α-Copaene<br>Germacren D<br>Chrysanthenyle acetate | α-Pinene<br>β-Pinene<br>Limonene<br>σ-Terpinene<br>Chrysantenone<br>*cis*-Carveol<br>α-Cedrene<br>α-Curcumene<br>α-Bergamotene<br>Cedrol<br>1-Octen-3-ol<br>*cis*-Sabinene hydrate |

* Underlined value indicates the rejection of the null of convergence at P <0.05 level of significance, while values in bold refer to the non-rejection of this hypothesis.

**Figure captions**

**Fig. 1** Factorial plan distribution of essential oils components from non-standardized Pearson clustering analysis

**Fig. 2** Dendrogram representing relationships based on similarities between compounds of *A. annua* sourced from nine locations

**Fig. 3** Transition curves of clusters 1 to 4 across nine locations with magnification of clusters 2 to 4 (clusters' compositions, see Table 2).

Fig. 1

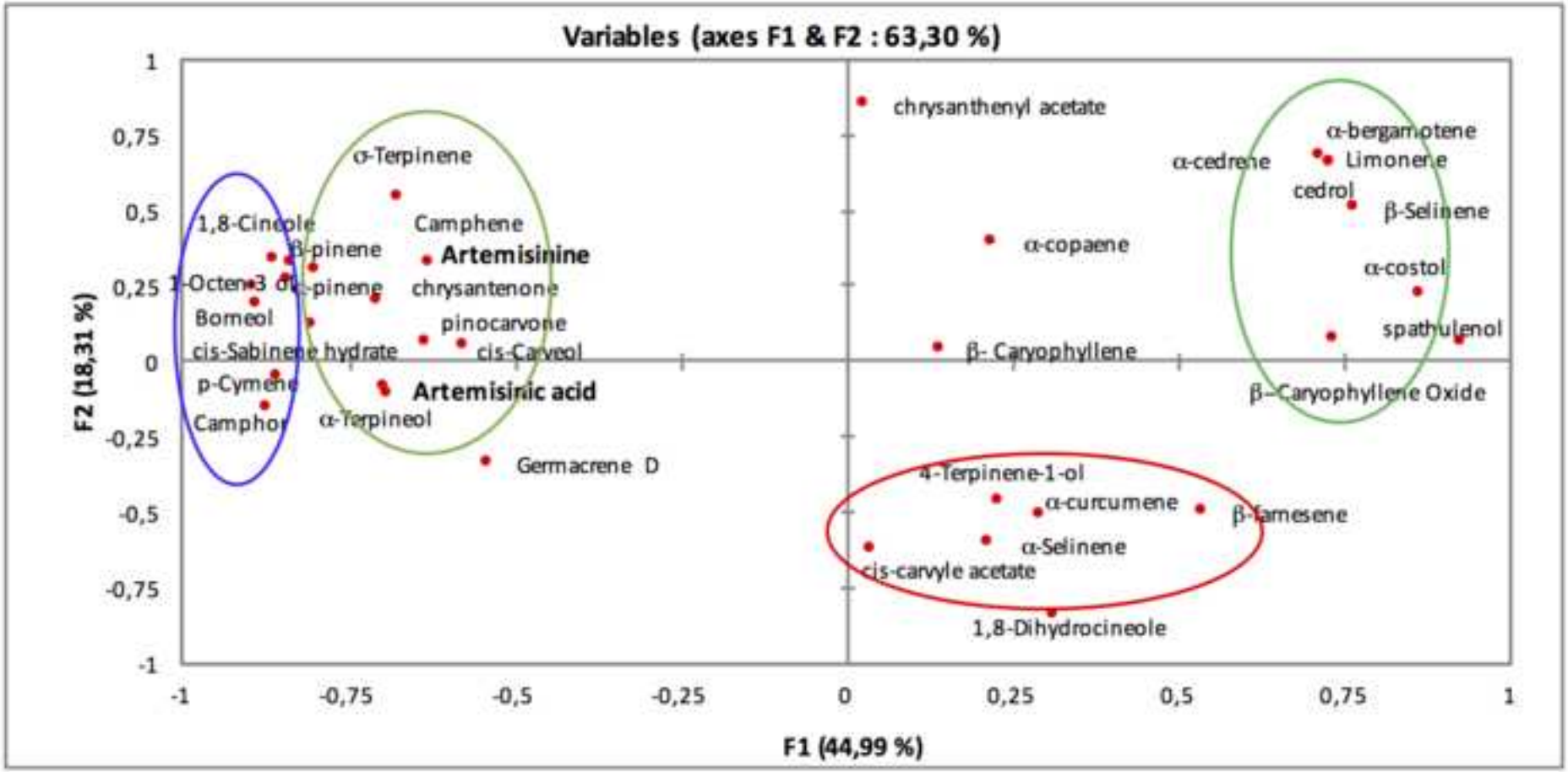

Variables (axes F1 & F2 : 63,30 %)
F2 (18,31 %)
F1 (44,99 %)
1
0,75
0,5
0,25
0
-0,25
-0,5
-0,75
-1
-1
-0,75
-0,5
-0,25
0
0,25
0,5
0,75
1
chrysanthenyl acetate
α-bergamotene
α-cedrene
Limonene
cedrol
β-Selinene
α-copaene
α-costol
spathulenol
β-Caryophyllene
β-Caryophyllene Oxide
α-Terpinene
1,8-Cineole
Camphene
β-pinene
Artemisinine
1-Octen-3-ol
α-pinene
chrysantenone
Borneol
pinocarvone
cis-Sabinene hydrate
cis-Carveol
p-Cymene
Artemisinic acid
Camphor
α-Terpineol
Germacrene D
4-Terpinene-1-ol
α-curcumene
β-famesene
α-Selinene
cis-carvyle acetate
1,8-Dihydrocineole

Fig. 2

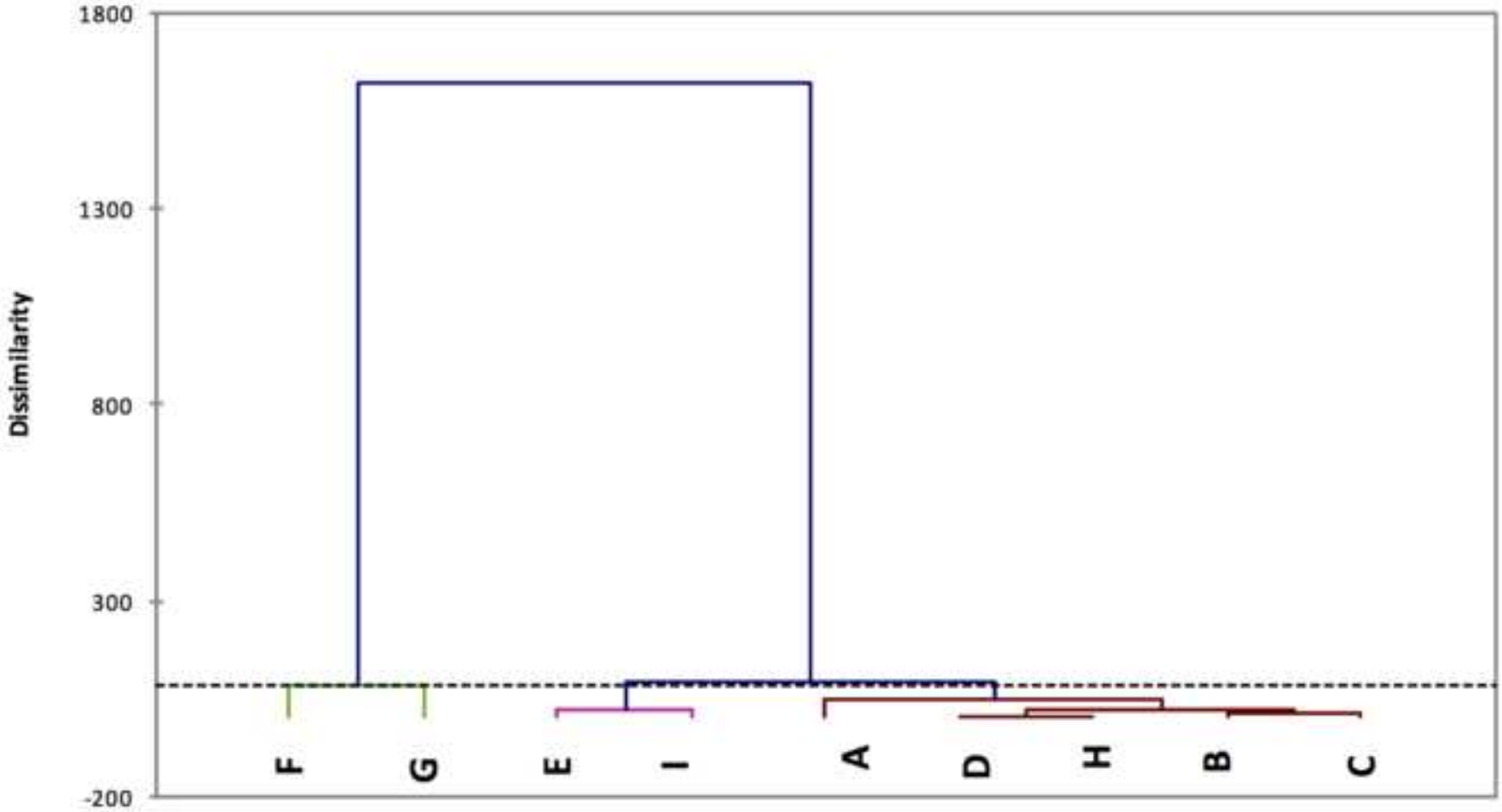
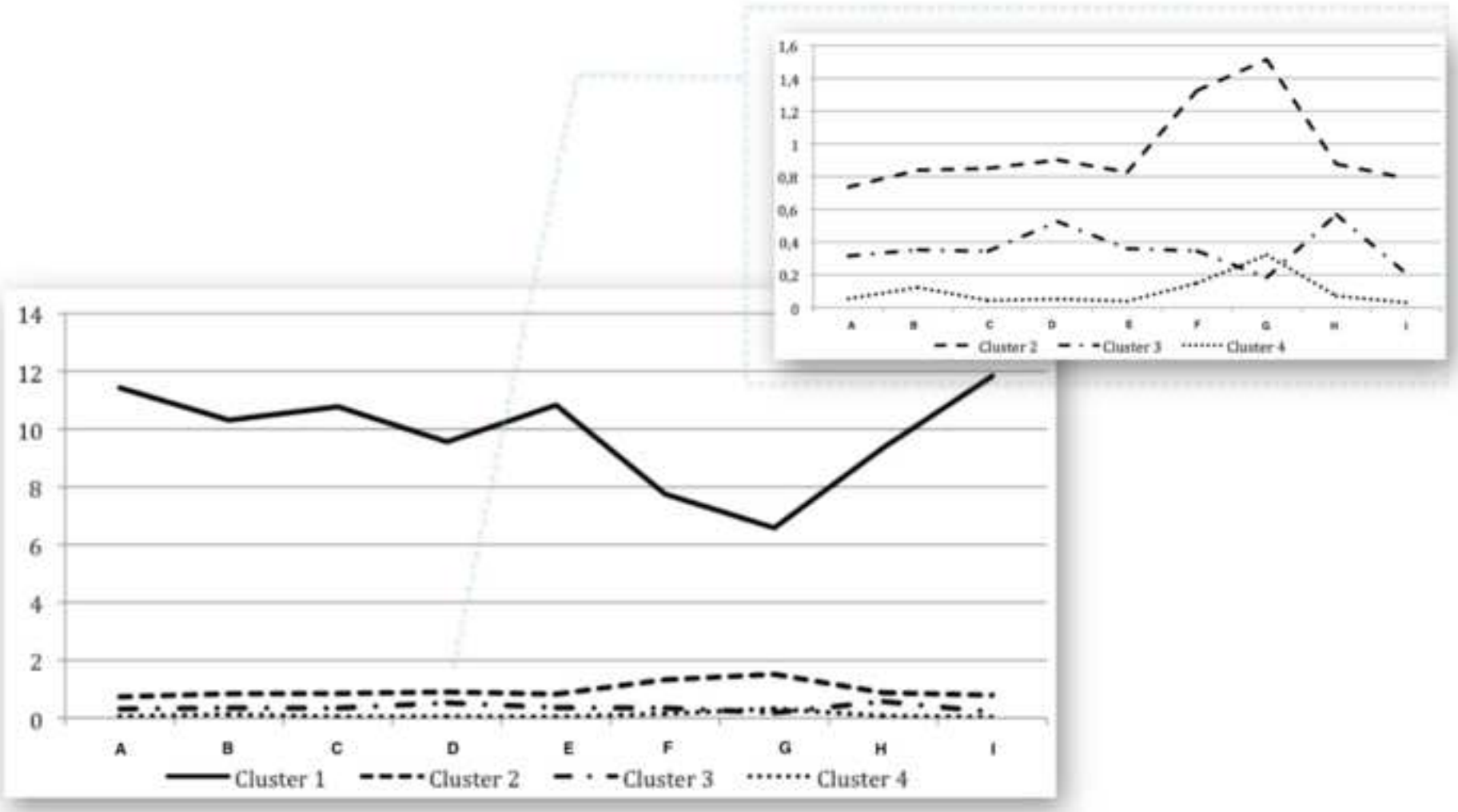


Fig. 3


14
12
10
8
6
4
2
0
A
B
C
D
E
F
G
H
I
Cluster 1
Cluster 2
Cluster 3
Cluster 4
1,6
1,4
1,2
1
0,8
0,6
0,4
0,2
0
A
B
C
D
E
F
G
H
I
Cluster 2
Cluster 3
Cluster 4