Combining ANOVA-PCA with POCHEMON to analyse micro-organism development in a polymicrobial environment☆

Brigitte P. Geurts [a], Anne H. Neerincx [b], Samuel Bertrand [c, d], Manja A.A.P. Leemans [a], Geert J. Postma [a], Jean-Luc Wolfender [d], Simona M. Cristescu [b], Lutgarde M.C. Buydens [a], Jeroen J. Jansen [a, *]

[a] Department of Analytical Chemistry, Institute for Molecules and Materials, Radboud University, Nijmegen, The Netherlands
[b] Department of Molecular and Laser Physics, Institute for Molecules and Materials, Radboud University, Nijmegen, The Netherlands
[c] Faculty of Pharmacy, University of Nantes, EA 2160-Mer Molécules Santé, 9 Rue Bias BP 53508, Nantes-cedex 1 44035, France
[d] School of Pharmaceutical Sciences, University of Geneva, University of Lausanne, CMU — Rue Michel Servet 1, 1211 Geneva 11, Switzerland

HIGHLIGHTS

- ANOVA-POCHEMON disentangles different information sources to study micro-organism development in a polymicrobial environment.
- It combines ANOVA with PCA of the isolated interspecies interaction-related chemistry in pathogen development.
- Two complementary co-culture studies show how it provides novel metabolic insight into interspecies interactions.

GRAPHICAL ABSTRACT
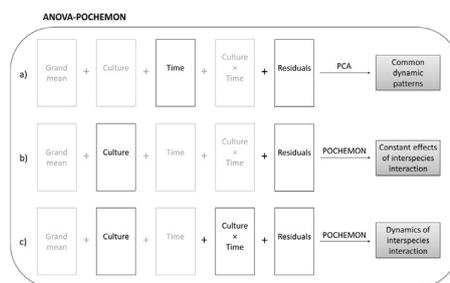
ABSTRACT

Revealing the biochemistry associated to micro-organismal interspecies interactions is highly relevant for many purposes. Each pathogen has a characteristic metabolic fingerprint that allows identification based on their unique multivariate biochemistry. When pathogen species come into mutual contact, their co-culture will display a chemistry that may be attributed both to mixing of the characteristic chemistries of the mono-cultures and to competition between the pathogens. Therefore, investigating pathogen development in a polymicrobial environment requires dedicated chemometric methods to untangle and focus upon these sources of variation. The multivariate data analysis method Projected Orthogonalised Chemical Encounter Monitoring (POCHEMON) is dedicated to highlight metabolites characteristic for the interaction of two micro-organisms in co-culture. However, this approach is currently limited to a single time-point, while development of polymicrobial interactions may be highly dynamic. A well-known multivariate implementation of Analysis of Variance (ANOVA) uses Principal Component Analysis (ANOVA-PCA). This allows the overall dynamics to be separated from the pathogen-specific chemistry to analyse the contributions of both aspects separately. For this reason, we propose to integrate ANOVA-PCA with the POCHEMON approach to disentangle the pathogen dynamics and the specific biochemistry in interspecies interactions. Two complementary case studies show great potential for both liquid and gas

chromatography - mass spectrometry to reveal novel information on chemistry specific to interspecies interaction during pathogen development.

1. Introduction

The interaction between different micro-organisms is important in many scientific fields. It may for one be a serious health problem: in patients with cystic fibrosis (CF), respiratory co-infections can lead to a higher exacerbation and hospitalization rate compared to patients only infected with one pathogen [1]. On the other hand, the unique biochemistry of co-occurring micro-organism may also be a way to enhance chemical diversity for drug discovery [2]. Co-occurring micro-organisms may also influence water quality [3,4] and be explicitly used in industrial fermentation processes [5,6].

Co-occurrence of micro-organisms can lead to interaction between the species. Interaction-related Metabolites may be 1) *de novo* produced, or 2) upregulated, or downregulated compared to the metabolites that the individual species produce [2,7]. These metabolite changes of interspecies interaction can be either beneficial or detrimental for both or for one of the species and, if there is any, to their human host (e.g. a human with respiratory infections) [8,9].

The complexity of the microbiome makes studying the interaction between different species *in vivo* a very challenging task. The metabolite production by pathogens can be different in the presence of other pathogens, and it may also be highly dynamic with regards to pathogen growth [7,10,11]. Therefore, *in vitro* studies are necessary to understand these complex biochemical interactions. Microorganism co-cultures (multiple microorganism species grown within a single confined environment) can be used to study how pathogens develop over time *in vitro*, as well as how they behave in close proximity of other micro-organisms [12]. The *de novo* produced compounds may exhibit interesting biological activities, such as antimicrobial and anticancer activities [2]. This makes microorganism co-cultures a promising approach to discover new natural bioactive compounds that can be used e.g. for medicinal purposes [7].

Detecting the induction of metabolite biosynthesis in microorganism co-culture requires sensitive metabolomic techniques mainly based on mass spectrometry [13]. Both liquid and gas chromatography coupled to mass spectrometry (LC-MS, GC-MS) provide efficient determination of metabolites produced by the pathogen(s) under study. Data analysis to find those metabolites characteristic for interspecies interaction is often done in a univariate manner [7,14,15]. However, a pathogen can often not be identified based on one characteristic metabolite, and a multivariate metabolite pattern is then required [16]. The metabolites are produced at different rates in different stages of the infection [7], which may provide invaluable information on the interaction dynamics. These patterns might be obscured by other natural variability in the data, such that a generic data analysis method may not detect them.

Dedicated chemometric methods may provide a comprehensive overview of the involved metabolites. Methods used for co-culture studies include Principal Component Analysis (PCA) [13], Analysis of Variance (ANOVA) [17,18], Self-Organizing Maps (SOM) [19], and multivariate Discriminant Analysis [10,13]. Although these methods provide insight in which aspects of the metabolic profiles are co-culture specific, they do not discriminate between the two different sources of co-culture biochemistry, i.e. mixing and

interspecies interaction. This means that the biochemistry related to interspecies interaction remains convoluted. Recently, we presented Projected Ortogonalized CHemical Encounter MONitoring (POCHEMON) to specifically highlight these metabolic alteration in co-culture [7]. However, the dynamics of pathogen development in co-culture cannot be directly assessed with POCHEMON or any other of the above mentioned methods.

Analysis of Variance can be used to separate the data into contributions related to different factors of variation in the data and their interactions [20,21]. Multivariate Analysis of Variance (MANOVA) is the extension of ANOVA to multiple independent variables, which has several disadvantages making it less applicable [21]. Several of these drawbacks may be overcome by regularization, involving an additional meta-parameter [22]. Several other multivariate implementations of ANOVA exist, which vary in the way the effect matrices are analysed. The most widely used methods are ANOVA-Simultaneous Component Analysis (ASCA) [23,24], and ANOVA-PCA [20,21]. In ASCA, PCA is applied directly onto each effect matrix. In ANOVA-PCA, PCA is applied to the sum of an effect matrix and the matrix of residuals. Other methods have been developed to perform PCA on biologically more relevant partitions than those obtained from 'standard' ANOVA models, such as Principal Response Curves [25] and SMART analysis [26], that fit within a generic framework that combines ANOVA and PCA [27]. Also alternatives for PCA, used within the ANOVA framework have been described such as Parallel Factor Analysis (PARAFASCA) [28] and Target Projection (ANOVA-TP) [29].

We propose the combination of ANOVA-PCA with POCHEMON for dedicated analysis of dynamic co-culture studies. This strategy allows for the extraction of three types of information:

1) information on the <u>dynamic</u> patterns common to both pathogens and to their co-culture,
2) information on the <u>constitutive</u> effect of interspecies interaction on pathogen metabolism, present at all stages of infection, and
3) information on the interspecies <u>interaction dynamics.</u>

We demonstrate this strategy on two complementary time-resolved microbial co-culture studies: an LC-MS study on *Aspergillus clavatus* and *Fusarium* sp. at four different time points at day level, and a GC-MS study of *Pseudomonas aeruginosa* and *Aspergillus fumigatus* at three different time points at hour level. The LC-MS study involves a fungus-fungus interaction where the metabolites are detected in the growth medium. Since the method is destructive, each time point measured involves different culture samples. In contrast, the GC-MS study involves a bacterium-fungus interaction where volatile metabolites are detected in the culture headspace such that the same samples may be followed over time. To assess the added value of the information from ANOVA-POCHEMON, we compare its results with its two constituent methods POCHEMON and ANOVA-PCA.

2. Theory

2.1. PCA

In PCA, a data matrix **X** is decomposed into a score matrix **T** and

a loading matrix $\mathbf{P}^T$ that capture the essential patterns in $\mathbf{X}$ [30]. Linear combinations of the original variables in $\mathbf{X}$ make up the new variables, called Principal Components (PCs). The scores hold the essential information of the data expressed on these PCs, whereas the loadings contain the relationship between the PCs and the original variables:

$$\mathbf{X} = \mathbf{T}\mathbf{P}^T + \mathbf{E} \qquad (1)$$

where $\mathbf{X}$ of dimensions $(I \times J)$ contains the data analysed on I replicates for J metabolite features; $\mathbf{T}$ of dimensions $(I \times R)$ contain the scores on the R PCs; $\mathbf{P}$ of dimensions $(J \times R)$ contain the corresponding loadings; and $\mathbf{E}$ is the matrix of residuals. The first PC is the direction that explains the most variation in the data; the second PC is the direction orthogonal to PC 1 that explains then the most variation, etc.

Principal Component Analysis is a well-established method to visualize variation in the data [30,31], and has successfully been applied to dynamic microorganism culture experiments before [11,32]. However, it is not possible to make a distinction between chemistry that is purely a mixture of the two mono-cultures, and chemistry caused by interaction between the two mono-cultures in a PCA model (*de novo* production, up- and/or downregulation of compounds), since PCA describes all variation in the data indiscriminately. both the mixing and the interaction chemistry are captured together in the same scores and loadings, and cannot be evaluated independently. Furthermore, collectively analysing multiple time points provides models that convolute the dynamic and consistent chemical variability, hampering the interpretation of models from time-resolved experiment [33]. Although local models of each time-point do not entangle this dynamic variability, these cannot be directly quantitatively compared because the loading basis of the different time points will vary, which also hampers their interpretability [34]. Therefore, PCA has shortcomings both to highlight 1) chemistry specific to interspecies interaction, or 2) to observe dynamic patterns in the data.

2.2. POCHEMON

POCHEMON can achieve the first of these aspects, highlight chemistry specific to interspecies interaction, by introducing two sequentially fitted PCA models [7]. The first PCA model of POCHEMON is called the 'Mixing model' and consists of a PCA on the mono-culture replicates of both species m_1 and m_2:

$$\begin{bmatrix} \mathbf{X}_{m_1} \\ \mathbf{X}_{m_2} \end{bmatrix} = \begin{bmatrix} \mathbf{T}_{\text{mix},m_1} \\ \mathbf{T}_{\text{mix},m_2} \end{bmatrix} \mathbf{P}_{\text{mix}}^T + \begin{bmatrix} \mathbf{E}_{\text{mix},m_1} \\ \mathbf{E}_{\text{mix},m_2} \end{bmatrix} \qquad (2)$$

where $\mathbf{X}_m$ of dimensions $(I_m \times J)$ contains the mono-culture data of species m, analysed on I_m replicates for J metabolite features; $\mathbf{T}_{\text{mix,m}}$ of dimensions $(I_m \times R_{mix})$ contain the mono-culture scores for species m on the R_{mix} Mixing PCs; $\mathbf{P}_{\text{mix}}$ of dimensions $(J \times R_{mix})$ contain the corresponding loadings; and matrix $\mathbf{E}_{\text{mix},m}$ of dimensions $(\sum_{m=1}^{M} I_m \times J)$ contains the mono-culture residuals. The Mixing scores show how much both species resemble each other through the separation of the scores into species-specific clusters. The variability among replicates of each mono-culture is revealed in the spread within the scores of each species on these Mixing components. Mixing scores for the co-culture replicates are obtained from the orthogonal projection of the co-culture data onto the Mixing loadings:

$$\mathbf{T}_{\text{mix},c} = \mathbf{X}_c \mathbf{P}_{\text{mix}} \qquad (3)$$

where $\mathbf{X}_c$ of dimensions $(I_c \times J)$ contains the co-culture data of I_c co-culture replicates (subscript c indicating co-culture data); and matrix $\mathbf{T}_{\text{mix},c}$ of dimensions $(I_c \times R_{mix})$ contains the Mixing scores of each co-culture replicate. These Mixing scores of the co-culture replicates express the composition of each co-culture replicate as a mixture of the metabolites in both separate species. Therefore, each of these Mixing scores is expected to be located between the mono-culture scores of both species.

The residuals of the co-culture projections are called the 'Mixing residuals', and they contain the information specifically related to interspecies interaction:

$$\mathbf{E}_{\text{mix},c} = \mathbf{X}_c - \mathbf{T}_{\text{mix},c}\mathbf{P}_{\text{mix}}^T \qquad (4)$$

The information in these residuals is then extracted by a second PCA model, the 'Competition model':

$$\mathbf{E}_{\text{mix},c} = \mathbf{T}_{\text{comp},c}\mathbf{P}_{\text{comp}}^T + \mathbf{E}_{\text{comp},c} \qquad (5)$$

where $\mathbf{T}_{\text{comp},c}$ of dimensions $(I_c \times R_{comp})$ contains the Competition scores; matrix $\mathbf{P}_{\text{comp}}$ of dimensions $(J \times R_{comp})$ the corresponding loadings; and matrix $\mathbf{E}_{\text{comp},c}$ of dimensions $(I_c \times J)$ contains the residuals of this model.

Orthogonal projection of the mono-culture residuals $\mathbf{E}_{\text{mix},m}$ onto the Competition model provides a 'baseline' or benchmark that expresses the natural variation among mono-culture replicates, against which the co-culture scores can be evaluated:

$$\begin{bmatrix} \mathbf{T}_{\text{comp},m_1} \\ \mathbf{T}_{\text{comp},m_2} \end{bmatrix} = \begin{bmatrix} \mathbf{E}_{\text{mix},m_1} \\ \mathbf{E}_{\text{mix},m_2} \end{bmatrix} \mathbf{P}_{\text{comp}} \qquad (6)$$

where $\mathbf{T}_{\text{comp},m}$ contains the mono-culture scores on the Competition loadings of dimensions $(\sum_{m=1}^{M} I_m \times J)$. The scores $\mathbf{T}_{\text{comp},m}$ are expected to surround the origin of the Competition model, and the Competition scores are only considered to contain co-culture specific information when they lie outside the benchmark of the mono-culture projections. In short, POCHEMON is able to separate mixing chemistry from interaction chemistry, but is unsuitable to explore and describe dynamic patterns.

2.3. ANOVA-PCA

Analysis of Variance in combination with PCA is one of the most commonly used models for data with a multilevel structure [21]. The original data matrix $\mathbf{X}$, containing all cultures and time points, may be decomposed into the sum of a series of sub-matrices, where each sub-matrix characterizes a factor of the experimental design. The residual matrix is then added back to each of these effect matrices, and PCA is performed on each of them separately to obtain the scores and loadings matrices for each variation source [35]. In time-resolved pathogen studies, this means that the data can be partitioned into sub-matrices according to the study design, as depicted in Fig. 1 [21].

The ANOVA model for this design with two factors of interest (Culture and Time) can be formulated as in Eq. (7). In this model the measured chromatogram $\mathbf{x}_{ijk}$ is assumed to be the result of the added effects of the factors Culture and Time over culture $i = 1, ..., I$ (being mono-culture 1, mono-culture 2 or the co-culture), time point $j = 1, ..., J$ and replicate $k = 1, ..., K$:

$$\mathbf{x}_{ijk} = \mu + \mathbf{x}_{\text{Culture},i} + \mathbf{x}_{\text{Time},j} + \mathbf{x}_{\text{Culture} \times \text{Time},ij} + \mathbf{x}_{\text{E},ijk} \qquad (7)$$

The effect of the interaction between Culture and Time, $\mathbf{x}_{\text{Culture} \times \text{Time},ij}$, is also incorporated in the ANOVA model shown here. The effects are added to a general mean expression value μ,
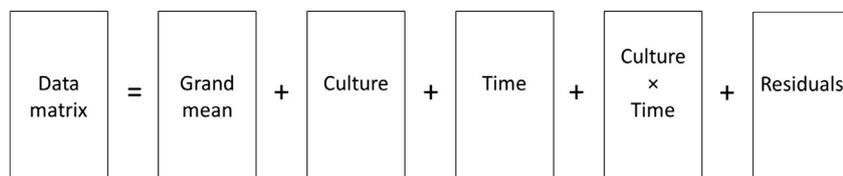
Fig. 1. A schematic of the decomposition of a data matrix of a time-resolved co-culture study into matrices that describe the sources of variation as the average, the variation of Culture type, Time, their interaction and the residuals.

and the remaining variation, the subject-specific effect, is captured in the error term $\mathbf{x}_{E,\ ijk}$. In ANOVA-PCA, the data matrix $\mathbf{X}$ ($IJK \times L$) is first decomposed into effect matrices according to the model in Equation (7) as:

$$\mathbf{X} = \mathbf{1}\mathbf{m}^{\mathrm{T}} + \mathbf{X}_{Culture} + \mathbf{X}_{Time} + \mathbf{X}_{Culture \times Time} + \mathbf{X}_{E} \qquad (8)$$

where $\mathbf{1}$ ($IJK \times 1$) consists of ones, $\mathbf{m}^{\mathrm{T}}$ ($1 \times L$) contains the means of the L variables computed across all IJK observations; $\mathbf{X}_{Culture}$ and $\mathbf{X}_{Time}$ hold the level means for the factors Culture and Time, respectively; $\mathbf{X}_{Culture \times Time}$ the interaction term for those two factors; and $\mathbf{X}_{E}$ the subject-specific effects (the residual matrix).

When the same samples are measured at all time points, the replicate variation can be separated as a third factor of repeated measurements. Because each sample belongs to only one of the cultures, this factor is nested within the Culture factor. This would lead to the extension of Equation (8) with the matrix $\mathbf{X}_{Replicate_{Culture}}$. Also an interaction term between Replicate and Time can be added to this equation. Including this Replicate factor imposes a repeated measures structure onto the ANOVA decomposition. For simplicity we omit these factors in the following procedures, but they can be analysed analogously to the others.

In ANOVA-PCA, PCA is performed after addition of the residual matrix to an effect matrix. This in contrary to ASCA, where the residual matrix is only projected into the PCA model on an effect matrix [35]. Although several comparison studies have shown that ASCA has some favorable properties [36,37], we have selected ANOVA-PCA here for reasons that will be pointed out in section 2.4. With the decomposition as described by Equation (8), we made three different PCA models: one on $\mathbf{X}_{Culture} + \mathbf{X}_{E}$ to visualize the overall effect of Culture, the second on $\mathbf{X}_{Time} + \mathbf{X}_{E}$ to visualize the general effect of Time, and a third to visualize the interaction between Culture and Time. This last model shows the culture-dependent time trends. Since $\mathbf{X}_{Culture \times Time}$ sums to zero for every combination of Culture and Time, the resulting PCA model contains very little biologically interesting information, as both the absolute and relative values of the individual values in this matrix can only be interpreted in terms of non-intuitive ANOVA constraints. For this reason, the culture factor and the relevant interaction were analysed combined in the original ASCA study and in many others that followed. The term $\mathbf{X}_{Culture \times Time}$ can still be used to assess the significance of this interaction, as described later in section 3.3. Following [28], the interaction term can be visualized for interpretational purposes by applying PCA to $\mathbf{X}_{Culture} + \mathbf{X}_{Culture \times Time} + \mathbf{X}_{E}$, identical to the contribution analysed in ASCA models with the residual matrix added. This model then describes all variation related to Culture and Time simultaneously. These three ANOVA-PCA models are described by Eqs. (9)–(11), respectively. Note that by removing the matrix of means from the data, the data is automatically mean-centered.

$$[\mathbf{X}_{Culture} + \mathbf{X}_{E}] = \mathbf{T}_{Culture}\mathbf{P}_{Culture}^{\mathrm{T}} + \mathbf{E}_{Culture} \qquad (9)$$

$$[\mathbf{X}_{Time} + \mathbf{X}_{E}] = \mathbf{T}_{Time}\mathbf{P}_{Time}^{\mathrm{T}} + \mathbf{E}_{Time} \qquad (10)$$

$$[\mathbf{X}_{Culture} + \mathbf{X}_{Culture \times Time} + \mathbf{X}_{E}] = \mathbf{T}_{Interaction}\mathbf{P}_{Interaction}^{\mathrm{T}} + \mathbf{E}_{Interaction} \qquad (11)$$

ANOVA is suitable to distinguish culture effects from dynamic patterns to allow separate analysis. Analysing the factor Time with PCA allows visualization and interpretation of dynamic patterns. However, as follows from section 2.1, analysis of the Culture factor with PCA does not provide information on the nature of the chemistry distinctive for co-cultures because PCA cannot distinguish mixing chemistry from chemistry related to interspecies interaction.

2.4. ANOVA-POCHEMON

Factor and interaction matrices from ANOVA may be analysed with multivariate methods other than PCA. The application of PLS on the effect matrices followed by target projection (ANOVA-TP) is suggested as an alternative to PCA [29], and the combination of ANOVA with PARAFAC (PARAFASCA) has shown great promise to analyse multilevel multiway data [28]. However, both generalizations are unfit to analyse the dynamic variability of main interest in co-culture studies.

Instead of performing PCA on the matrix $\mathbf{X}_{Culture} + \mathbf{X}_{E}$ from ANOVA, we can also apply POCHEMON to this matrix (ANOVA-POCHEMON). Application of ANOVA (following Eq. (8)) enables the separate analysis of the factors Time and Culture, and POCHEMON (following Eq. (2)) on the Culture effect allows the distinction between mixing chemistry and chemistry related to interspecies interaction. The Mixing model of POCHEMON on the Culture effect can be described by Eq. (12). Note that this involves a PCA model built on mono-culture data, as described by Eq. (2), on the addition matrix of $\mathbf{X}_{Culture}$ and $\mathbf{X}_{E}$. For simplicity reasons, the resulting scores, loadings, and residuals are subscripted only with 'Culture'.

$$\begin{bmatrix} \mathbf{X}_{Culture,m_1} + \mathbf{X}_{E,m_1} \\ \mathbf{X}_{Culture,m_2} + \mathbf{X}_{E,m_2} \end{bmatrix} = \begin{bmatrix} \mathbf{T}_{Culture,\mathrm{mix},m_1} \\ \mathbf{T}_{Culture,\mathrm{mix},m_2} \end{bmatrix} \mathbf{P}_{Culture,\mathrm{mix}}^{\mathrm{T}} + \begin{bmatrix} \mathbf{E}_{Culture,\mathrm{mix},m_1} \\ \mathbf{E}_{Culture,\mathrm{mix},m_2} \end{bmatrix} \qquad (12)$$

This is equal to Eq. (9) except that the co-culture data is not used to build this model, only the mono-culture data is.

The other steps of POCHEMON can applied to the culture factor in an analogous way following Eqs. (3)–(6):

$$\mathbf{T}_{Culture,\mathrm{mix},c} = [\mathbf{X}_{Culture,\ c} + \mathbf{X}_{E,c}]\mathbf{P}_{Culture,\mathrm{mix}} \qquad (13)$$

$$\mathbf{E}_{Culture,\mathrm{mix},c} = [\mathbf{X}_{Culture,c} + \mathbf{X}_{E,c}] - \mathbf{T}_{Culture,\mathrm{mix},c}\mathbf{P}_{Culture,\mathrm{mix}}^{\mathrm{T}} \qquad (14)$$

$$\mathbf{E}_{Culture,\mathrm{mix},c} = \mathbf{T}_{Culture,\mathrm{comp},c}\mathbf{P}^{\mathrm{T}}_{Culture,\mathrm{comp}} + \mathbf{E}_{Culture,\mathrm{comp}} \qquad (15)$$

$$\begin{bmatrix} \mathbf{T}_{Culture,\mathrm{comp},m_1} \\ \mathbf{T}_{Culture,\mathrm{comp},m_2} \end{bmatrix} = \begin{bmatrix} \mathbf{E}_{Culture,\mathrm{mix},m_1} \\ \mathbf{E}_{Culture,\mathrm{mix},m_2} \end{bmatrix} \mathbf{P}_{Culture,\mathrm{comp}} \qquad (16)$$

Equation (13) calculates the co-culture scores of the Mixing model, $\mathbf{T}_{Culture,\ \mathrm{mix},c}$; and Eq. (14) describes the co-culture residuals of this model, $\mathbf{E}_{Culture,\mathrm{mix},c}$. Equation (15) describes the Competition scores $\mathbf{T}_{Culture,\mathrm{comp},c}$ and loadings $\mathbf{P}^{\mathrm{T}}_{Culture,\mathrm{comp}}$ from the co-culture residuals, and Eq. (16) projects the mono-culture samples onto the Competition model.

This clarifies our reason for using ANOVA-PCA rather than ASCA. In ASCA, the co-culture scores of each sample are identical, since ASCA does not add the individual variation $\mathbf{X}_{\mathrm{E},c}$ to $\mathbf{X}_{Culture,\ c}$, but projects them into the model to represent variability. This implies that the co-culture residuals for every sample are also identical, such that a second PCA model based only on these residuals is obsolete.

The factor Time is independent of culture type, and for that reason this factor will be analysed with PCA as described in section 2.3. On the other hand, the Culture-Time interaction, expressed as $\mathbf{X}_{Culture} + \mathbf{X}_{Culture \times Time} + \mathbf{X}_{\mathrm{E}}$, can be analysed with POCHEMON (analogous to Eqs. (12)–(16)) to establish the chemistry of interspecies interaction specific for each separate time point.

Summarizing, ANOVA-POCHEMON entails four major parts:

1) Decomposition of the data matrix into effect matrices Culture, Time, and their interaction; following Eq. (8)
2) PCA on the factor Time (defined as $\mathbf{X}_{Time} + \mathbf{X}_{\mathrm{E}}$); following Eq. (10),
3) POCHEMON on the factor Culture (defined as $\mathbf{X}_{Culture} + \mathbf{X}_{\mathrm{E}}$); following Eqs. (13)–(16), and
4) POCHEMON on the Culture-Time interaction (defined as $\mathbf{X}_{Culture} + \mathbf{X}_{Culture \times Time} + \mathbf{X}_{\mathrm{E}}$); following an adaptation of Eqs. (13)–(16) for interaction.

In the introduction we mentioned three types of information present in dynamic co-culture data that can be extracted by ANOVA-POCHEMON. The workflow of ANOVA-POCHEMON is schematically depicted in Fig. 2, showing parts 2–4. Fig. 2a shows part 2 of the workflow: it analyses the dynamic patterns common to both pathogens and to their co-culture. Fig. 2b depicts part 3 of ANOVA-POCHEMON: it extracts the constitutive effect of interspecies interaction on pathogen metabolism, present at all time points. Fig. 2c shows part 4: it allows focus on the dynamics of interspecies interaction.

3. Materials and methods

3.1. LC-MS data

The experimental data used here were obtained from a previously described co-culture experiment [7]. It contained monocultures of fungal species *A. clavatus* (Sin141 – isolated from soil) and *Fusarium* sp. (PS54743 – isolated from a blood sample), stored in the database of Agroscope ACW (Swiss Federal Research Station, Wädenswil, Route de Duillier, P.O. Box 1012, CH-1260 Nyon, Switzerland, http://mycoscope.bcis.ch/). Both fungal strains were cultivated or co-cultivated in 12-well plates with 2 mL of potato dextrose agar (PDA, Difco). Strains were inoculated by placing 2-mm agar plugs of fungal pre-cultures in the centre or on the opposite sides of well for mono- and co-culture respectively. The cultures were incubated at 21 °C for 2, 4, 7 or 9 days. Both mono-

and the co-cultures were analysed with six replicates (n = 6). The lyophilized agar with mycelium was extracted by dichloromethane−methanol−water (64:36:8) under sonication for 20 min, following a previously published protocol [12]. Finally, the extracts were dried and the extreme non-polar constituents were removed by reversed-phase solid phase extraction.

Samples were fingerprinted in randomized order by the UHPLC-TOFMS platform (Acquity UPLC system coupled to a Micromass-LCT Premier Time-of-Flight mass spectrometer by a electrospray interface, Waters, Baden-Daettwil, Switzerland), with a 50 mm × 1 mm i.d. 1.7 μm Acquity BEH C$_{18}$ UPLC column (Waters, Baden-Daettwil, Switzerland) and a water/acetonitrile gradient that has been detailed in the full protocol [7]. Data were acquired in positive and negative ionization mode in a m/z of 100−1000. The LC-MS chromatograms were converted to peak lists of features with their corresponding retention time, m/z and peak area in each of the analysed samples using MZmine2 [38]. Data from the positive and negative ionization mode were combined by concatenation. The resulting peak areas were log-transformed data ($\log(x+1)$).

Significant metabolite features were putatively annotated by dereplication, which is the action of identification of observed features [39] (Level 2 annotation according to the Metabolomic Standard Initiative − MSI [40]). The molecular formula was determined based on the high mass-accuracy (maximum 10 ppm difference between the theoretical and measured m/z value), supported by isotopic pattern matching and heuristic filtering [41]. They were matched against metabolites in the Dictionary of Natural Products [42] that are produced by *Aspergillus* sp. and *Fusarium* sp.

3.2. GC-MS data

Cultures of *P. aeruginosa* strain ATCC 27853 and *A. fumigatus* clinical isolate AZN 8196 were obtained as described earlier [10]. To summarize, *P. aeruginosa* was inoculated into 50 mL Brain Heart Infusion (BHI) broth (Mediaproducts BV, The Netherlands) in an initial concentration of approximately 5×10^6 colony forming units (CFU)/mL. *Aspergillus fumigatus* was cultured on Sabouraud dextrose agar supplemented with 0.02% chloramphenicol, and the conidia were suspended into 50 mL of BHI broth plus 0.01% Tween 80 (Boom BV, The Netherlands) to a concentration of approximately 2.6×10^5 CFU/mL. Co-cultures (cultures with both *A. fumigatus* and *P. aeruginosa*) were obtained by preparing a culture of *A. fumigatus* as described above. *Pseudomonas aeruginosa* (5×10^6 CFU/mL) was manually added fifteen hours after the inoculation of *A. fumigatus*.

For sampling the headspace of the cultures, a setup was used as previously reported [43]. In short, the cultures were placed in 250 mL Erlenmeyer flasks. Each flask was closed using a glass stopper which contained two Teflon open-close valves acting as the inlet and outlet. The headspaces of the cultures were constantly flushed with catalysed compressed air. Headspace samples (3.5 L) were taken at 16, 24 and 48 h after inoculation of the first pathogen, *A. fumigatus*, by connecting a glass tube filled with Tenax TA$^{®}$ (Shimadzu, Japan) to the outlet of the stopper for 60 min. Each experiment was performed in 12 replicates (n = 12). The headspace samples were analysed using thermal desorption (TD20) coupled to QP2010 Ultra GC-MS (Shimadzu, Japan). Only compounds that were present in at least 50% of the replicates per culture were included for analysis, and expressed by the Total Ion Current (TIC). The resulting peak area was log-transformed data ($\log(x+1)$). Compounds were putatively identified based on 80% minimal similarity of the MS spectra compared to the National Institute of Standards and Technology (NIST) libraries NIST08 and NIST08s.
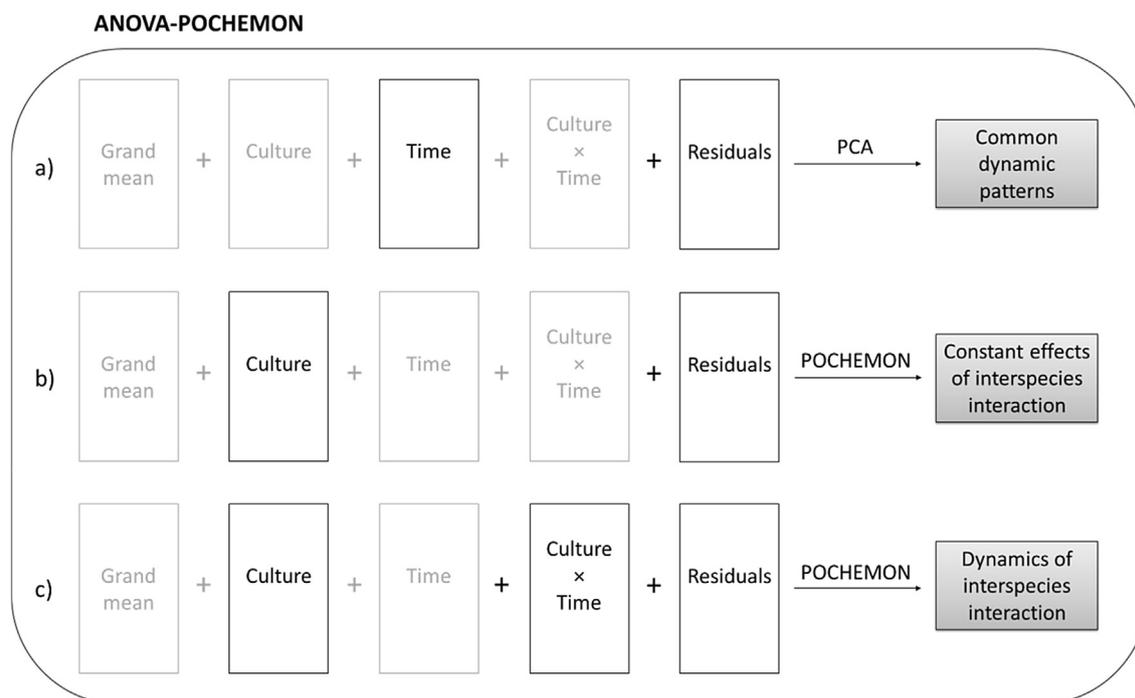
ANOVA-POCHEMON

Fig. 2. A schematic overview of ANOVA-POCHEMON, which entails a) PCA on the Time effect to extract common dynamic patterns, b) POCHEMON on the Culture effect to extract constant effects of interspecies interaction, and c) POCHEMON on the Culture-Time interaction to extract information on the dynamics of interspecies interaction.

3.3. Data analysis

We estimated the significance of the experimental factors in ANOVA using permutation tests to evaluate the summed value from all univariate sums of squares (SSQ) for the variables in each effect matrix [44]. Significance was defined by a sufficiently low fraction of permuted effect matrices with a total SSQ larger than the total SSQ of the observed effect matrix.

In the GC-MS data the same samples were measured at different time points, meaning that a third factor was added to the ANOVA-model describing the Replicate variation, which is nested within the Culture factor.

For POCHEMON and ANOVA-POCHEMON, the dimensionality of the competition model has been established from a scree plot, combined with an evaluation of the interpretability of the competition scores. The monoculture benchmark served as a validation here, as it indicated whether variability in the co-culture scores was also present in the projections of the monocultures on the competition model.

For POCHEMON and ANOVA-POCHEMON, the most important metabolic features were determined by resampling validation based on Jansen et al. (2014) [7]. For the LC-MS data, random replicates for each time point were selected to construct Mixing and Competition models on. This resampling procedure was repeated 6400 times to average out the effect of randomly selecting mono-culture replicates of each time point, and the final rank products were determined from these repetitions. For the GC-MS data, single replicates including all time point were selected systematically to construct Mixing and Competition models on. For each resampling, the Competition loadings were ranked, and an overall rank product of each feature was determined across all resampling realizations. To establish the number of important metabolic features, a rank product threshold was set based on a steep increase in rank product when including more features. If no steep increase could be detected, the threshold was put at ten

features.

All algorithms used in this paper are coded and executed in MATLAB version 8.3.0 (Mathworks, Natick MA).

4. Results and discussion

In this section the results will be presented and discussed for the LC-MS and GC-MS data subsequently. For each data set, results of POCHEMON, ANOVA-PCA and ANOVA-POCHEMON will be presented in separate subsections, where the strengths and weaknesses of each method as described in section 2 are pointed out. This will be done more elaborately in the LC-MS section, but the observations translate to the GC-MS example. For all Mixing and Competition models, two PCs were chosen to be optimal.

4.1. LC-MS data

Results for separate PCA and POCHEMON models for each individual time point are given in Supplementary Figs. 1 and 2. It is very difficult to interpret any information about dynamic processes from these figures, since it requires comparing different PCA models. A straight-forward comparison of the scores of different models is only possible if they are expressed on the same loading base [34]. However, it is clear that the cultures differ at several (but not all) time points. The fact that the co-culture samples in Supplementary Fig. 1 do not score directly in between of the mono-cultures, indicates that they contain more than only a mixture chemistry of the two mono-cultures. Additionally, the POCHEMON Competition models in Supplementary Fig. 2 indicate both a dynamic effect and an effect of interspecies interaction in the data. Therefore, both POCHEMON (to study the interspecies interaction) and ANOVA-PCA (to investigate the dynamic variability) may bring added value to the analysis. Moreover, the combination between these two methods may highlight the dynamics of interspecies interaction.
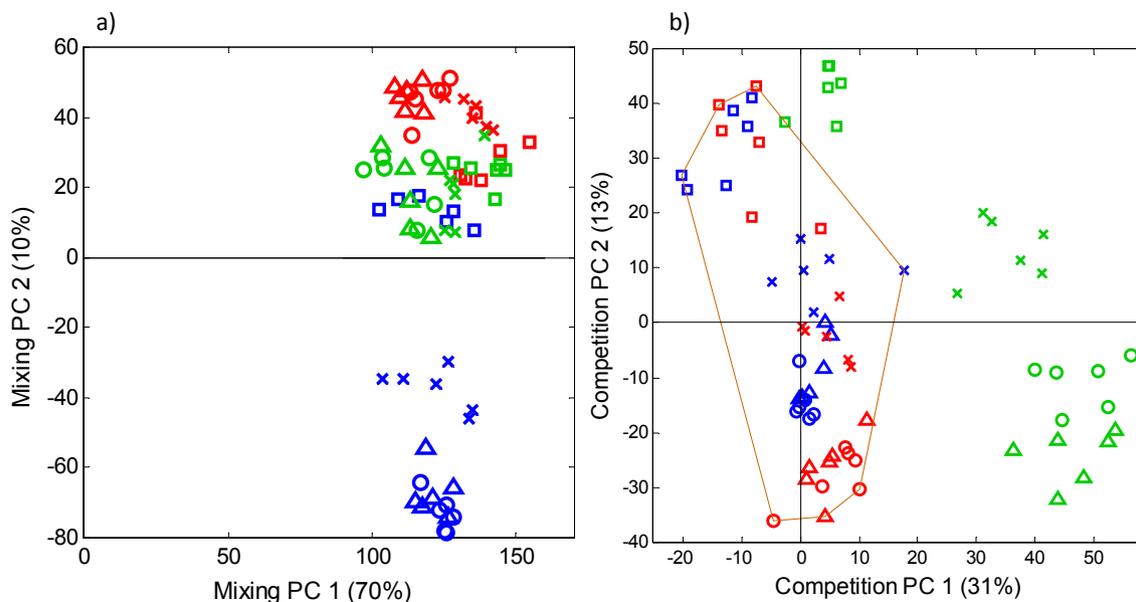
Fig. 3. POCHEMON a) mixing model, and b) competition model on data of all time points. *Fusarium* sp. samples are indicated in blue, *A. clavatus* in red and their co-culture in green. Symbols reflect time after inoculation, squares (□) representing 2 days, crosses (×) 4 days, circles (○) 7 days and triangles (Δ) 9 days. The orange zone corresponded to the mono-cultures benchmark in the competition model. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

4.1.1. POCHEMON

Fig. 3 shows both the Mixing model and the Competition model from POCHEMON (on all time points). The Mixing model (Fig. 3a) focuses on the mono-culture chemistry of both pathogen cultures, and shows a great diversity among the *Fusarium* sp. replicates. Two days after inoculation the *Fusarium* sp. mono-culture samples were very similar to the co-culture samples and even to the *A. clavatus* mono-culture samples. This can be explained by slow growth of the fungi, not producing many compounds after two days yet leading to insignificant differences between the LC-MS profiles [7]. However, after four days the *Fusarium* sp. mono-cultures scored very differently from the *A. clavatus* on the second Mixing Component. This difference increased on day seven, and remained unchanged until day nine. This complies with the trends in the decrease in glucose content reported [7]: the confined space was almost fully saturated after seven days, leaving no more room for development. In the mono-cultures of *A. clavatus* and in the co-cultures the dynamic variability was described mainly in Mixing Component 2, where samples scores increased over time. A similar dynamic effect can be observed in the Competition model (Fig. 3b). In this model, samples clustered based on time after inoculation. The co-culture samples scored outside of the benchmark formed by the mono-cultures, indicating the presence of chemistry specific to interspecies interaction. However, the dynamic pattern makes analysis of the loadings and identification of the compounds involved difficult, because this information is convoluted with the interspecies interaction in both sub-models. This is why these sources of variation should be separated.

4.1.2. ANOVA-PCA

Separating the culture variation from the dynamic variability by ANOVA allows analysis of the cultures and, separately, of their common dynamic patterns. The factors Culture, Time and their interaction all contributed significantly to the model ($p < 0.001$). Fig. 4 shows the results of ANOVA-PCA for this LC-MS data set. Fig. 4a shows a perfect separation between the three culture types (mono-culture 1, mono-culture 2 and their co-culture) on the first two PCs. There is no dynamic pattern detectable, proving the

concept of the ANOVA separation of factors. Additionally, the co-culture scores are not in between the two mono-cultures, indicating that there is more going on than only a mixture of two co-cultures. However, this model is not dedicated to highlight important compounds for interspecies interaction. Fig. 4b shows that the dynamic variability (the factor Time) can be described by the first PC of this PCA model. The scores increased from day two until day seven on PC1. Day nine had similar scores as day seven, indicating that there was less growth in the last two days of the experiment. The ten loadings of the most important metabolic features describing dynamic variability (the largest loadings on PC1) are listed in Table 1. The fact that the Time effect can be described by one PC only means that the responsible features are relevant for the dynamics at all four time points. Fig. 4c shows the Culture-Time interaction effect. The three cultures all showed very similar behavior on day two, but then developed each in different directions of the score plot. This figure provides a dedicated view on the dynamics specific for each of the cultures. However, this model focuses on variation, meaning that the most dynamic culture is considered most important to the model. Moreover, with this analysis it is obscured which part of the co-culture dynamics is specific for interspecies interaction, and which part is a mixture from mono-culture dynamics.

We showed in this section that PCA of the factor Time provides information on the dynamic patterns common to both pathogens and their co-culture. This is step 2 of the workflow and corresponds with Fig. 2a from section 2.4. Additionally, we showed that the ANOVA decomposition is effective in removing the influence of this dynamic variability from the data describing the Culture effect. However, PCA is not a suitable method to focus on interspecies interaction.

4.1.3. ANOVA-POCHEMON

As shown in Fig. 2, ANOVA-POCHEMON entails three parts. The results of the first part, PCA on the Time factor from ANOVA (Fig. 2a), has already been described in section 4.1.2. In this section, the results of ANOVA-POCHMON of both the Culture factor (Fig. 2b) and the Culture-Time interaction (Fig. 2c) are discussed.
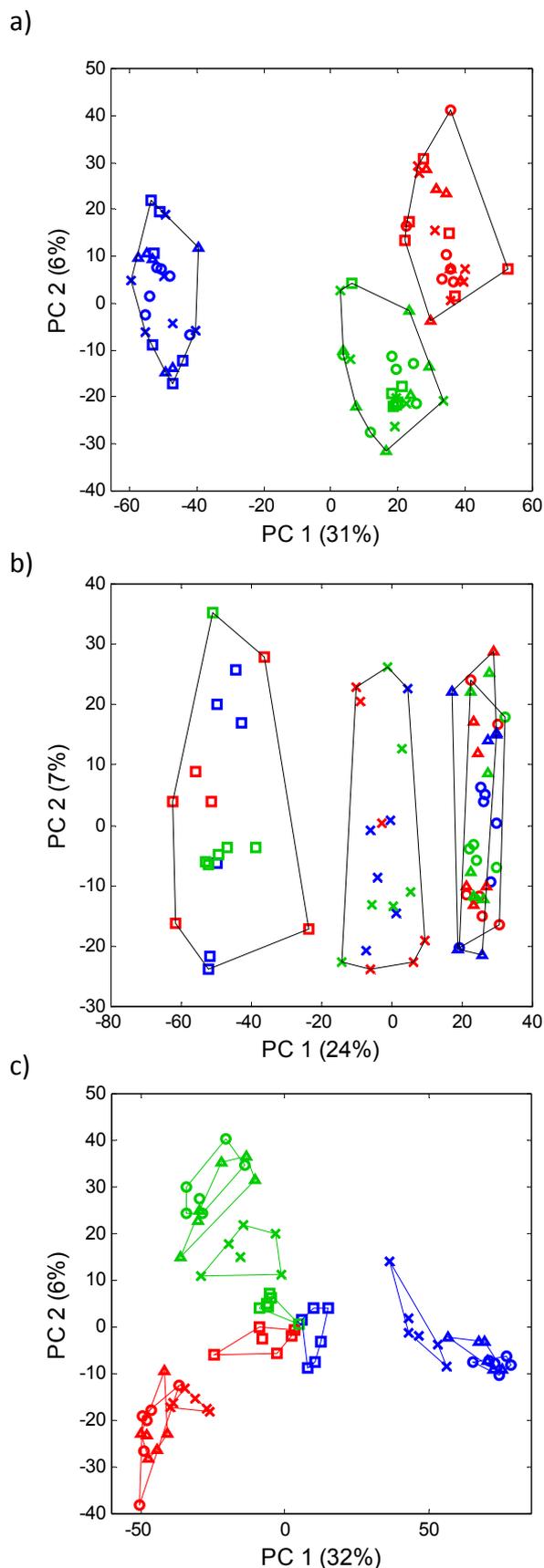
a)

b)

c)

Fig. 4. ANOVA-PCA score plot of a) the Culture factor, convex hulls indicating the different culture types; b) the Time factor, convex hulls indicating the different time points; and c) the Time-Culture interaction, convex hulls indicating the different

We combined the benefits of POCHEMON with those of ANOVA-PCA by performing POCHEMON on the relevant ANOVA factors and interactions, i.e. those that involve Culture, to highlight the chemistry specific to interspecies interaction within the co-culture. The sub-models in Fig. 5 provide information on the effect of interspecies interaction on fungal behavior. This allows us to answer the same questions as with POCHEMON in its original application, without interfering from general dynamic aspects:

1) How do the mono-cultures differ chemically (at all time points)?
2) How do the co-cultures differ from the mono-cultures (at all time points)?

Additionally, we can answer similar questions regarding the development of the cultures:

1) How does the development of the mono-cultures differ chemically (over all time points)?
2) How does the development of the co-cultures differ from that of the mono-cultures (over all time points)?

The Mixing model for the Culture factor shows that the samples of the two mono-cultures were well separated (Fig. 5a). In the ANOVA-PCA model the co-cultures did not score between the two mono-cultures, whereas in the mixture model of ANOVA-POCHEMON the co-culture scores indeed represent a linear combination of the chemistry between both species. The projected co-culture samples scored very close to *A. clavatus*, showing that the chemistry of this fungus dominates the mixture throughout the experiment. Note that whereas the POCHEMON Mixing model (Fig. 3) is not centered, the ANOVA-POCHEMON sub-model is. This is the consequence of subtraction of the grand mean in ANOVA, that we did to focus on independently explaining the different sources of variation. Because the main focus of this work is to isolate the interspecies interaction, we will not further analyse the Mixing model in depth here and focus only on the second question listed above.

The Competition model for the Culture factor (Fig. 5b) shows that the scores of all co-culture samples exceeded those of the mono-culture benchmark, indicating that all samples exhibited chemistry specific to interspecies interaction. Positive scores on the first Competition component indicated the consistent information across all co-culture replicates. The variation amongst the co-culture replicates was captured in the second Competition component. The loadings of the first Competition PC directly reflect the importance of each compound to the distinct, consistent biochemistry of the co-cultures, which likely relates to interspecies interaction. The resampling validation revealed fifteen important metabolite features, ions present in the UHPLC-TOFMS analyses. These are listed in Table 2 and represented by arrows in Fig. 5b. Eight metabolite features (printed in bold in Table 2) are the main contributors to interspecies interaction, since their loadings are largest on the first Competition component.

Five of these eight metabolite features have been indicated as *de novo* produced, upregulated or downregulated in co-culture by an earlier univariate analysis on this data [7]. Metabolite feature #5, NI287.043@1.43 (the notation corresponds to an ion detected using the negative ionization mode that had an *m/z* of 287.043 at a

culture types at different time points. *Fusarium* sp. samples are indicated in blue, *A. clavatus* in red and their co-culture in green. Symbols reflect time after inoculation, squares (□) representing 2 days, crosses (×) 4 days, circles (○) 7 days and triangles (Δ) 9 days. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

Table 1
Highlighted metabolic features for overall dynamics LC-MS data.

Metabolite feature#	Ionization mode	Retention time (min)	m/z (Da)	Putative IDs
8	Positive	1.51	361.124	8-Keto-15-acetyl-3*; 7,8,15-Scirpenetetrol*; 3-Acetyl-3,7,15-trihydroxy-8-scirpenone*
2	Positive	2.69	279.229	Unidentified
9	Negative	0.95	234.061	Unidentified
7	Negative	1.45	285.043	Unidentified
6	Negative	1.51	221.0811	3,4-Dihydro-6,8-dihydroxy-3,4,5-trimethyl-1H-2-benzopyran-1-one**; Asparvenone**; 7,8-Dihydro-7,8-dihydroxy-3,5,7-trimethyl-6H-2-benzopyran-6-one**; Parvulenone**
3	Negative	2.02	311.222	5,8-Dihydroxy-9,12-octadecadienoic acid**
5	Negative	2.14	309.205	Unidentified
10	Negative	2.35	311.222	5,8-Dihydroxy-9,12-octadecadienoic acid**
4	Negative	2.70	564.330	Unidentified
1	Negative	3.50	559.473	Unidentified

The last column indicates putative identities of the features, where one asterisk (*) indicates that this compound is known to be produced by *Fusarium* spp, and a double asterisk (**) indicates compounds known to be produced by *Aspergillus* spp.

retention time of 1.43 min), is suspected to come from the same metabolite as feature #3 (PI289.069@1.46), which has been highlighted as upregulated and produced on a longer time span in co-culture [7]. The univariate analysis highlighted metabolic features that were produced in different concentrations in co-culture than in the mono-cultures. The ANOVA-POCHEMON model of the Culture factor focuses only on those combinations of features that are consistently different for all time points. Since many of the features highlighted by univariate analysis were only different for some of the time points, not all these features may be selected by this ANOVA-POCHEMON submodel. Additionally, the strength of multivariate analysis is that it can reveal features that are only important in combination with others. The compounds NI285.043@1.45 and NI341.196@1.75 were not highlighted by univariate analysis and are thus novel, multivariate discoveries from ANOVA-POCHEMON. None of those two features were successfully dereplicated, showing the great potential of co-cultivation for the induction of unreported compounds. Using ANOVA-POCHEMON, these compounds are for the first time highlighted as specific for the interspecies interaction between *A. clavatus* and *Fusarium* sp. and they can be investigated further in future research.

Also the Culture-Time interaction contains contributions of both mixture and interspecies interaction that can be disentangled with POCHEMON. As explained in section 2.3, the Culture-Time interaction factor is not easily interpretable when expressed only as $\mathbf{X}_{Culture \times Time} + \mathbf{X}_E$. This is shown in Supplementary Fig. 3. Supplementary Fig. 3a shows the Mixing model, in which the scores of each factor setting sum to zero due to the ANOVA decomposition. This makes the assessment of trends regarding culture-specific interspecies interaction very challenging. For this reason, we have expressed the Culture-Time interaction as $\mathbf{X}_{Culture} + \mathbf{X}_{Culture \times Time} + \mathbf{X}_E$. Fig. 5c shows the Mixing model for the Culture-Time interaction when expressed as such, describing the dynamic variability in the mixing chemistry of the co-cultures. There is a greater contribution to this plot for the time points four, seven and nine days from the mono-cultures, the co-culture scores lying in between the mono-culture scores, on Mixing component 1.

In the Competition model (Fig. 5d) the scores on Competition component 1 increased over time. A sub-cluster of samples appeared in with high scores on Competition component 2, consisting of samples from the last two time points.

The ten metabolic features most important for this dynamic pattern of interspecies interaction are listed in Table 3 and indicated in Fig. 5d with arrows. The loadings of this model were highly dissimilar to those of the ANOVA-PCA on the Culture-Time interaction. This underlines again that ANOVA-PCA is not able to focus

on the dynamic information specifically related to interspecies interaction. All of these ten metabolic features were also highlighted for global interspecies interaction (Table 2). The sub-cluster in the top of Fig. 5d is characterized by lower concentrations of these features compared to other samples of the same time points. A possible reason for the different scores of this sub-cluster might be that their growth is confined by the limited space available. However, we do not have information about the size of the fungal colonies in the different replicates to confirm this. Three out of the ten features were most important for the culture-time interaction, and are highlighted in bold in Table 3. These three features were also highlighted in Table 2 for global interspecies interaction, and they were also highlighted by univariate analysis as *de novo* produced, upregulated or downregulated in co-culture in a previous study [7]. The results from ANOVA-POCHEMON provide the novel knowledge that these features are also specifically important to describe the multivariate development of interspecies interaction over time.

We have shown that POCHEMON on the ANOVA Culture factor provides information on the effect of interspecies interaction on fungal behavior. For the Culture-Time interaction, POCHEMON allows abstraction of the dynamics of interspecies interaction, leading to more insight in the metabolic features involved. By combining ANOVA-PCA with POCHEMON according to the workflow shown in Fig. 2, we have developed one method to extract.

a) information on the constitutive effect of interspecies interaction on pathogen metabolism, present at all time points (Fig. 4b);
b) information on the dynamic patterns common to both pathogens and to their co-culture (Fig. 5b); and
c) information on the dynamics of interspecies interaction (Fig. 5d).

4.2. GC-MS data

Results of PCA and POCHEMON per time point results can be found in Supplementary Figs. 4 and 5. The different cultures differ at several (but not all) time points, indicating the presence of both a dynamic and an interspecies effect in the data.

4.2.1. POCHEMON

Similar to our findings in the LC-MS study, application of POCHEMON to all samples for the GC-MS data of *P. aeruginosa* and *A. fumigatus* showed large differences between the measurement time points (Fig. 6). The mixing model (Fig. 6a) indicates a considerable spread within each mono-culture, and even overlap between the two mono-cultures. The co-cultures are spread widely
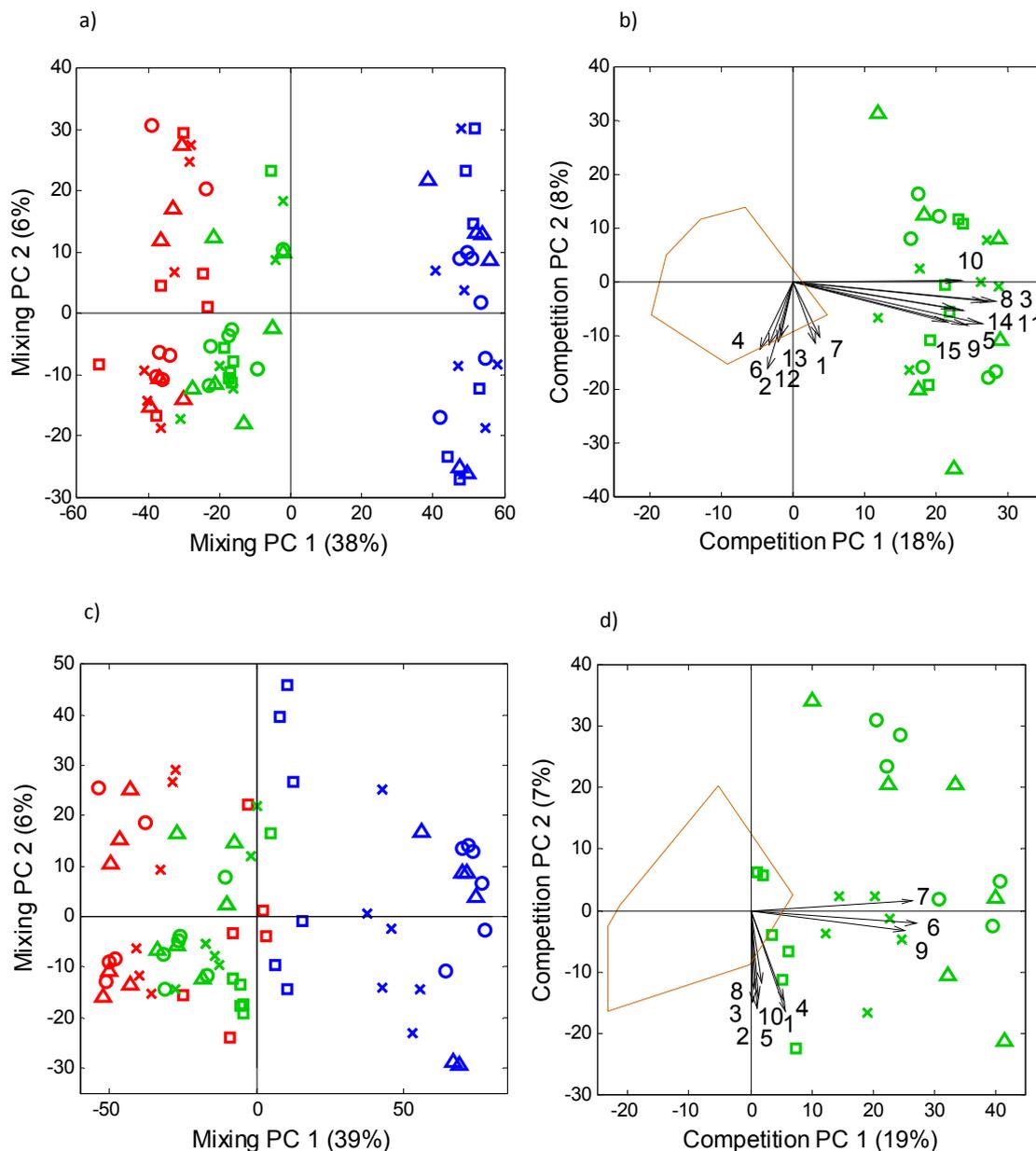
Fig. 5. ANOVA-POCHEMON a) mixing model for the Culture factor, b) competition model for the Culture factor where the arrows represent the loadings related to the metabolic features listed in Table 2, c) mixing model for the Time-Culture interaction, d) competition model for the Time-Culture interaction where the arrows represent the loadings related to the metabolic features listed in Table 3. *Fusarium* sp. samples are indicated in blue, *A. clavatus* in red and their co-culture in green. Symbols reflect time after inoculation, squares (□) representing 2 days, crosses (×) 4 days, circles (○) 7 days and triangles (△) 9 days. The orange zone corresponded to the mono-cultures benchmark in the competition model. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

amongst the mono-culture samples. There is also a clear influence of the time after inoculation visible, leading from 16 h at the bottom to 48 h at the top of the figure. This dynamic variability is also obvious in the competition model (Fig. 6b), were the co-culture samples clustered based on time after inoculation. These results underline the conclusions drawn for the previous example: POCHEMON on samples from different time points does not suffice to provide detailed information about interspecies interaction nor on development of the compounds over time. An additional source of information that POCHEMON cannot use is the repeated analysis of samples from the same individual, obtained at different time-points. Quantification of this Replicate effect by ANOVA decomposition may further focus the analysis upon the relevant sources of variation.

4.2.2. ANOVA-PCA

The ANOVA effects of Culture, Time, Replicate, and the interaction between Culture and Time (defined as $\mathbf{X}_{Culture} + \mathbf{X}_{Culture \times Time}$) were significant ($p = 0.036$ for Sample and $p < 0.001$ for the others). Fig. 7 shows the results of ANOVA-PCA. The mono-cultures in Fig. 7a are well separated, with the co-culture scores showing overlap with *A. fumigatus*. The difference between the culture types is mainly related to compounds with large loadings on the first principal component. The most important compounds have higher concentrations in the samples of *A. fumigatus*, and they are 8-nonen-2-one, 2-nonanone, 2-trideconene, 2-undecanone. These compounds were also present in the discriminative time-independent biomarker profile determined previously with this data, without the use of ANOVA [10]. This underlines the ability of

Table 2
Highlighted metabolic features for interspecies interaction LC-MS data.

Metabolite feature#	Ionization mode	Retention time (min)	*m/z* (Da)	Putative IDs
3	**Positive**	**1.46**	**289.069**	**Eurotinone**; Funalenone**; O-demethyl-7-methoxy-funalenone**; Anhydrofusarubin***
14	**Positive**	**1.85**	**286.070**	**Bostrycoidin***
15	**Positive**	**2.55**	**265.178**	***Unidentified***
12	Positive	4.09	861.582	*Unidentified*
2	Positive	4.10	859.571	*Unidentified*
6	Positive	4.12	861.589	*Unidentified*
13	Positive	4.25	822.705	*Unidentified*
1	Positive	4.30	394.336	*Unidentified*
4	Positive	4.32	427.370	*Unidentified*
7	Positive	4.33	450.352	*Unidentified*
9	**Negative**	**1.28**	**151.039**	**Gibepyrone F*; 2-Methoxy-6-methyl-1,4-benzoquinone**; Phenoxyacetic acid****
5	**Negative**	**1.43**	**287.043**	***Unidentified***
11	**Negative**	**1.44**	**205.087**	***Unidentified***
8	**Negative**	**1.45**	**285.043**	***Unidentified***
10	**Negative**	**1.75**	**341.196**	***Unidentified***

Bold rows indicate features specifically important for interspecies interaction. The last column indicates putative identities of the features, where one asterisk (*) indicates that this compound is known to be produced by *Fusarium* spp, and a double asterisk (**) indicates compounds known to be produced by *Aspergillus* spp.

Table 3
Highlighted metabolic features for culture-time interaction LC-MS data.

Metabolite feature#	Ionization mode	Retention time (min)	*m/z* (Da)	Putative IDs
7	**Positive**	**1.46**	**289.069**	**Eurotinone**; Funalenone**; O-demethyl-7-methoxy-funalenone**; Anhydrofusarubin***
8	Positive	4.09	861.582	*Unidentified*
2	Positive	4.10	859.571	*Unidentified*
3	Positive	4.12	861.589	*Unidentified*
10	Positive	4.25	822.705	*Unidentified*
1	Positive	4.30	394.336	*Unidentified*
5	Positive	4.32	427.370	*Unidentified*
4	Positive	4.33	450.352	*Unidentified*
9	**Negative**	**1.28**	**151.039**	**Gibepyrone F*; 2-Methoxy-6-methyl-1,4-benzoquinone**; Phenoxyacetic acid****
6	**Negative**	**1.43**	**287.043**	***Unidentified***

Bold rows indicate features specifically important for interspecies interaction. The last column indicates putative identities of the features, where one asterisk (*) indicates that this compound is known to be produced by *Fusarium* spp, and a double asterisk (**) indicates compounds known to be produced by *Aspergillus* spp.
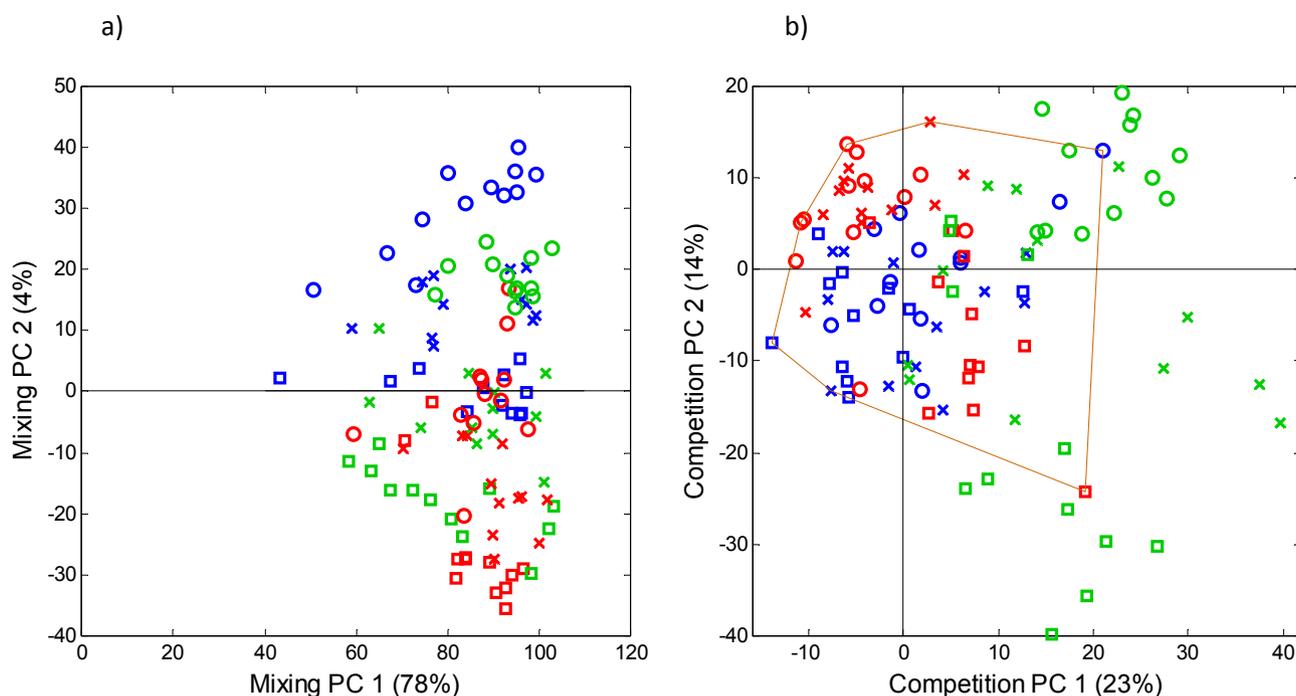

Fig. 6. POCHEMON a) mixing model, and b) competition model on data of all time points. *Pseudomonas aeruginosa* samples are indicated in blue, *A. fumigatus* in red and their co-culture in green. Symbols reflect time after inoculation, squares (□) representing 16 h, crosses (×) 24 h and circles (○) 48 h. The orange zone corresponded to the mono-cultures benchmark in the competition model. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)
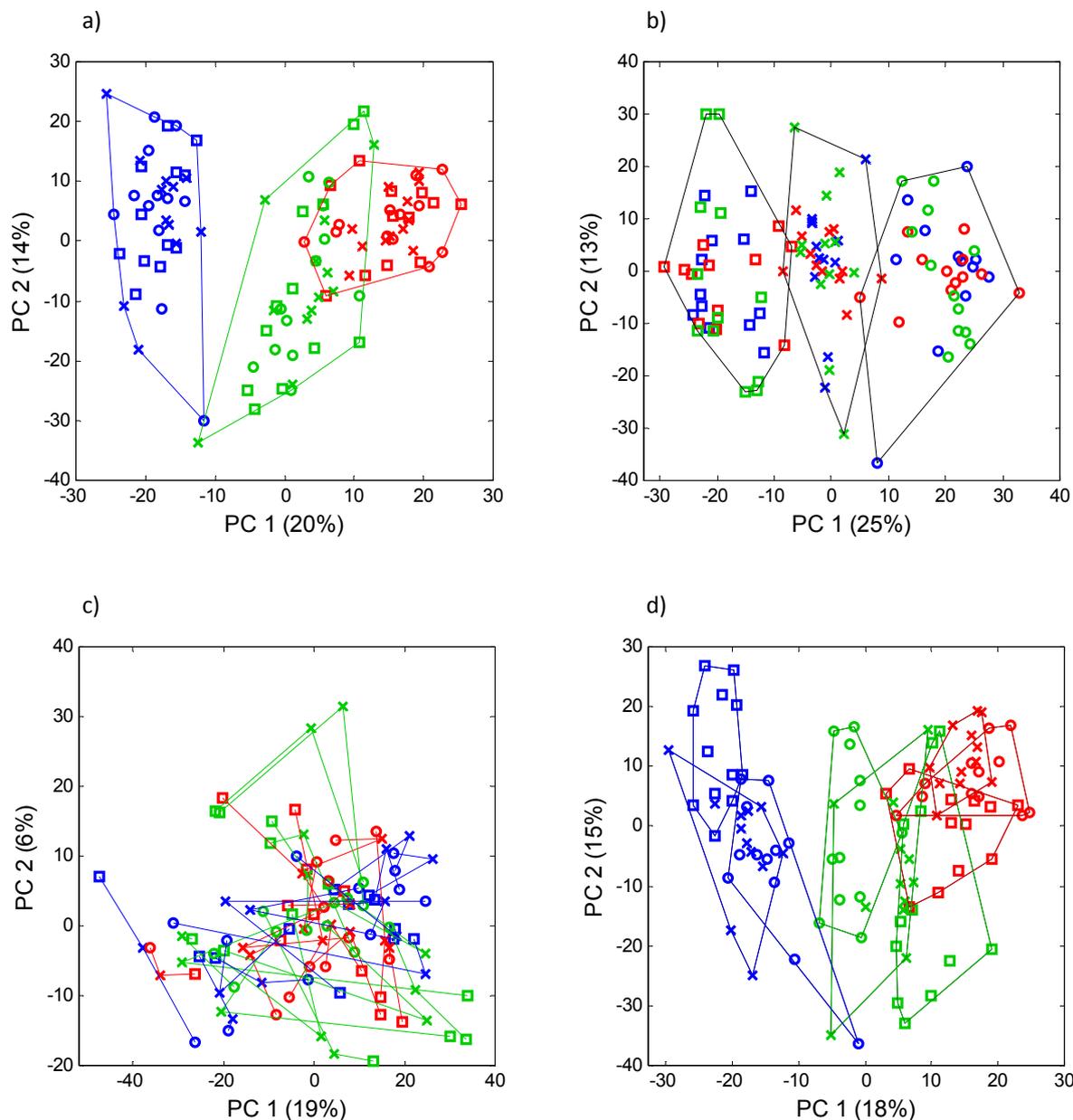
Fig. 7. ANOVA-PCA score plot of a) the Culture factor, convex hulls indicating the different culture types; b) the Time factor, convex hulls indicating the different time points; c) the Replicate factor, lines connecting measurements of the same replicate sample; and d) the Time-Culture interaction, convex hulls indicating the different culture types at different time points. *Pseudomonas aeruginosa* samples are indicated in blue, *A. fumigatus* in red and their co-culture in green. Symbols reflect time after inoculation, squares (□) representing 16 h, crosses (×) 24 h and circles (○) 48 h. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

ANOVA to separate different sources of variability. Although this model provides some insight in the differences between the mono-cultures, it does not highlight important metabolic features for interspecies competition.

Fig. 7b displays the Time effect. The scores show a clear dynamic pattern from the left of the graph to the right. The compounds that contribute mainly to this pattern are listed in Table 4. It is probable that several of these compounds that show a common development in both cultures and their co-culture are emitted by the growth medium, which is the same for all cultures.

The Replicate effect (i.e. variation within the individual replicates) is shown in Fig. 7c. There is no systematic variability within the scores of this factor. However, removal of this variation may still enhance the information content of other factors, as this removes

Table 4
Highlighted metabolic features for overall dynamics GC-MS data.

Metabolite feature#	Putative IDs
1	*Unidentified*
2	Dimethyl disulfide
3	Dimethyl trisulfide
4	Trimethylamine
5	Methyl isobutyl ketone
6	4-Methyl-3-penten-2-one
7	Methyl ester thiocyanic acid
8	2-Methyl-1-butanol
9	1,4-Pentadiene
10	2-Methyl-2-Butenal

any interference of individual variation within replicates.

Fig. 7d displays the Culture-Time interaction. Each culture shows a different development over time. The *P. aeruginosa* samples score high on the second PC in the beginning, and low at the end of the experiment. Samples from *A. fumigatus* and the co-culture show an opposite development, although less explicit.

This case study confirms our statement on ANOVA-PCA in section 4.1.2: Separate PCA analysis of the factor Time provides unique information on the common dynamic patterns in the data. Analysis of the Culture effect or the Culture Time interaction does not highlight information on interspecies interaction. These ANOVA effects should be analysed with a more dedicated tool. Additionally, we showed here that the ANOVA model is flexible to different designs by the inclusion of a factor Replicate, nested within Culture. This allowed removal of the disturbing individual variation.

4.2.3. ANOVA-POCHEMON

The results of the first part of ANOVA-POCHEMON, PCA on the Time factor from ANOVA (as schematically depicted in Fig. 2a), has already been described in section 4.1.2. In this section, the results of ANOVA-POCHMON of both the Culture factor (Fig. 2b) and the Culture-Time interaction (Fig. 2c) are discussed. The main difference with the LC-MS case study described above is that the ANOVA decomposition allowed also removal of a Replicate effect, because this experiment involved non-destructive measuring of the same samples at the different points.

The constitutive interspecies interaction was investigated by POCHEMON on the Culture effect from ANOVA. The Mixing model is depicted in Fig. 8a, and shows that the scores of the two mono-cultures were well separated. The projected co-culture samples fell in between both mono-cultures, although they scored more similar to *A. fumigatus* than to *P. aeruginosa*.

In the Competition model the majority of the co-culture samples exceeded the mono-culture benchmark, indicating that in most samples there was chemistry present specific to interspecies interaction. The positive score on the first Competition component indicated the consistent information across all co-culture replicates. The variation between the replicates was captured in the second component. Green lines connect measurements of the same replicate sample. This allows us to compare trajectories of each replicate. Some replicates show very similar effects of interspecies interactions, while others are very different. This means that ANOVA-POCHEMON can be used also to assess replication of observed responses between replicate experiments, and analyse how the highlighted metabolic features are influenced by the variation between the replicates. The resampling validation revealed important metabolite features, VOCs from the GC-MS analysis. The ten statistically most important compounds are listed in Table 5 and the corresponding loadings are indicated by arrows in Fig. 8b. In contrast to the LC-MS case study, most loadings are large on both axes. This indicates that the corresponding compounds are important for the general interspecies interaction as well as the variation among replicates.

The compounds 1-undecene, 2-ethyl-benzenamine, 1-propanol and 2-hexanone were upregulated upon interspecies interaction, as mono-culture replicates contain them in considerably lower amounts (data not shown). Metabolites 1-hydroxy-2-propanone and chloro-benzene were downregulated upon interspecies interaction. None of these features are *de novo* produced or eliminated upon interspecies interaction.

The dynamics of interspecies interactions were investigated by POCHEMON on the Culture-Time interaction. Supplementary Fig. 6 shows that also for this data set the Culture-Time interaction is not easily interpretable when expressed only as $X_{Culture \times Time} + X_E$. For this reason, Fig. 8c shows the Mixing model for the Culture-Time

interaction expressed as $X_{Culture} + X_{Culture \times Time} + X_E$, describing the dynamic variability in the mixing chemistry of the co-cultures. Fig. 8d models the interspecies interaction at each stage of infection. After 16 h the majority of samples did not express dynamic interspecies interaction, as they scored within the mono-culture benchmark. Since the second pathogen was added to the culture only after 15 h, this is as expected. Most samples collected after 24 and 48 h scored outside this mono-culture benchmark. The replicate samples can be followed in time by their individual trajectories plotted in the figure. Some samples have a trajectory moving from left to right over time, while others show an angled trajectory that can point both to the left and right of the figure. This reveals that the dynamic response to interspecies interaction is highly individual-specific. The eight most important metabolic features for the dynamics of the interspecies interaction for all replicates are listed in Table 6. These features show great overlap with those listed in Table 5. This indicates that the systematic differences between the cultures are considerably larger than the dynamic variability in these differences. The only compounds unique to Table 6 is 2-methyl-1-butanol.

ANOVA-POCHEMON is the first method to allow disentanglement of the different chemistries in dynamic co-cultures, and therefore there is no established benchmark to confirm our findings. A recent study by Briard et al. showed that the presence of dimethyl-sulfide, produced by *P. aeruginosa*, has a stimulatory effect on the growth of *A. fumigatus* [45]. However, we did not detect any dimethyl-sulfide in the *P. aeruginosa* headspace in our study and for this reason we cannot confirm these findings.

As mentioned in section 4.1.3, the Mixing model of POCHEMON is not mean-centered whereas the ANOVA-POCHEMON sub-model is. There is a specific merit to omit centering in this situation: unlike chemical mixtures, mixing micro-organisms may lead to 'dillution' of the sample when both species die. This may lead to a decrease in the total amount of molecules. In the non-centered, resulting cone-shaped mixing model, these co-cultures would be projected closer to the origin than monocultures, but still in between the cultures of both species. Centering convolutes such information in the model: in ANOVA-POCHEMON indeed such effects could be more difficult to observe.

Evaluation of the explained variance of the Competition model compared to the total variance in the data in POCHEMON is currently challenging. Further research is required to assess different figures of merit and to develop a comprehensive validation tool for POCHEMON. It follows that in ANOVA-POCHEMON, a numerical assessment of the significance of a variables' contribution to the Competition model would be even more challenging. However, the amount of variance explained by the model is not of primary importance, the specific structure of that variance is. Even 1−2% of 'variation' in the measured data would be relevant, as it may be attributed to a smaller contribution that is either isolated by the ANOVA-based ANOVA-PCA operations or by the POCHEMON step. What is mainly important here, is that this variation is present and is statistically validated by 1) check of reproducibility between co-culture replicates, and 2) comparison to the mono-culture benchmark.

5. Conclusion

In the present study a dedicated, novel approach for the analysis of dynamic chromatographic data coming from *in vitro* micro-organism co-culture has been presented. This approach combined the strengths of ANOVA-PCA and POCHEMON. For two case studies, we showed that ANOVA was suitable to separate the information in the data by different sources of variation, namely Time and Culture. PCA on the Time effect provided insight into which metabolites are
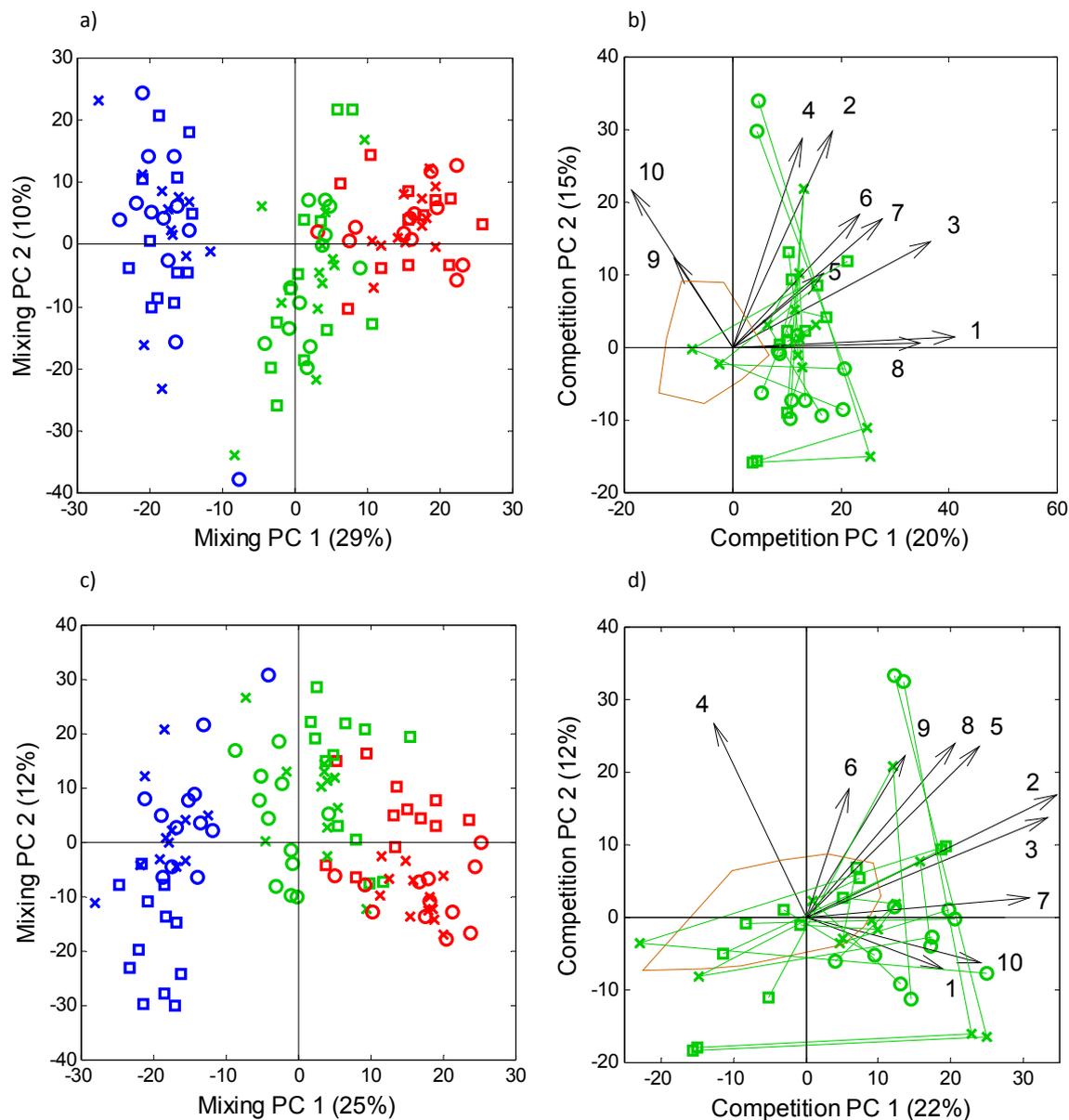
Fig. 8. ANOVA-POCHEMON a) mixing model for the Culture factor, b) competition model for the Culture factor where the arrows represent the loadings related to the metabolic features listed in Table 5, green lines connecting measurements of the same replicate sample, c) mixing model for the Time-Culture interaction, d) competition model for the Time-Culture interaction where the arrows represent the loadings related to the metabolic features listed in Table 6, green lines connecting measurements of the same replicate sample. *Pseudomonas aeruginosa* samples are indicated in blue, *A. fumigatus* in red and their co-culture in green. Symbols reflect time after inoculation, squares (□) representing 16 h, crosses (×) 24 h and circles (○) 48 h. The orange zone corresponded to the mono-cultures benchmark in the competition model. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

Table 5
Highlighted metabolic features for interspecies interaction GC-MS data.

Metabolite feature#	Putative IDs
1	1-Undecene
2	3-Methyl-2-butenal
3	2-Ethyl-benzenamine
4	2-Decanone
5	3-Hydroxy-2-butanone
6	3-Methyl-1H-pyrrole
7	1-Propanol
8	2-Hexanone
9	1-Hydroxy-2-propanone
10	Chloro-benzene

Table 6
Highlighted metabolic features for culture-time interaction.

Metabolite feature#	Putative IDs
1	1-Undecene
2	3-Hydroxy-2-butanone
3	1-Propanol
4	Chloro-benzene
5	3-Methyl-2-butenal
6	1-Hydroxy-2-propanone
7	2-Ethyl-benzenamine
8	2-Decanone
9	3-Methyl-1H-pyrrole
10	2-Methyl-1-butanol

involved in pathogen development, regardless of the nature of the pathogen. However, PCA was not able to untangle the complex metabolic profiles in the Culture effect, e.g. separate the mixed mono-culture biochemistries from a chemistry specific to their interspecies competition. The method POCHEMON was designed specifically for that purpose, but we showed that it was not suitable for dynamic data due to convolution with dynamic patterns. We demonstrated that application of POCHEMON on the Culture effect of ANOVA lead to the discovery of metabolites involved in inter-species interaction at all time points. Additionally, ANOVA-POCHEMON revealed how metabolite patterns influenced by interspecies interaction, changed over time.

In conclusion, ANOVA-POCHEMON provides a powerful tool for the simultaneous discovery of metabolite profiles 1) related to dynamic processes, 2) specific for interspecies competition, and 3) related to the dynamics of this interspecies interaction. The pre-sented case studies indicate that this method leads to novel metabolic insights that may improve diagnosis of co-infections, e.g. in the lungs of CF patients, or lead to the discovery of new natural bioactive lead medicinal compounds.

Acknowledgements

Appendix A. Supplementary data

Supplementary data related to this article can be found at http://dx.doi.org/10.1016/j.aca.2017.01.064.

References

[1] R. Amin, A. Dupuis, S.D. Aaron, F. Ratjen, The effect of chronic infection with *Aspergillus fumigatus* on lung function and hospitalization in patients with cystic fibrosis, CHEST J. 137 (1) (2010) 171–176.

[2] S. Bertrand, N. Bohni, S. Schnee, O. Schumpp, K. Gindro, J.-L. Wolfender, Metabolite induction via microorganism co-culture: a potential way to enhance chemical diversity for drug discovery, Biotechnol. Adv. 32 (6) (2014) 1180–1204.

[3] A.S. Beliaev, M.F. Romine, M. Serres, H.C. Bernstein, B.E. Linggi, L.M. Markillie, N.G. Isern, W.B. Chrisler, L.A. Kucek, E.A. Hill, G.E. Pinchuk, Inference of in-teractions in cyanobacterial–heterotrophic co-cultures via transcriptome sequencing, ISME J. 8 (11) (2014) 2243–2255.

[4] W. Manz, M. Wagner, R. Amann, K.H. Schleifer, In situ characterization of the microbial consortia active in two wastewater treatment plants, Water Res. 28 (8) (1994) 1715–1723.

[5] J.A. Narvhus, T.H. Gadaga, The role of interaction between yeasts and lactic acid bacteria in African fermented milks: a review, Int. J. Food Microbiol. 86 (1–2) (2003) 51–60.

[6] E.J. Smid, C. Lacroix, Microbe–microbe interactions in mixed culture food fermentations, Curr. Opin. Biotechnol. 24 (2) (2013) 148–154.

[7] S. Bertrand, A. Azzollini, O. Schumpp, N. Bohni, J. Schrenzel, M. Monod, K. Gindro, J.-L. Wolfender, Multi-well fungal co-culture for de novo metabo-lite-induction in time series studies based on untargeted metabolomics, Mol. BioSyst. 10 (9) (2014) 2289–2298.

[8] D.K. Morales, D.A. Hogan, Candida albicans interactions with bacteria in the context of human health and disease, PLoS Pathog. 6 (4) (2010).

[9] M.E. Shirtliff, B.M. Peters, M.A. Jabra-Rizk, Cross-kingdom interactions: candida albicans and bacteria, FEMS Microbiol. Lett. 299 (1) (2009) 1–8.

[10] A.H. Neerincx, B.P. Geurts, M.F.J. Habets, J.A. Booij, J. Van Loon, J.J. Jansen, L.M.C. Buydens, J. van Ingen, J.W. Mouton, F.J.M. Harren, R.A. Wevers, Iden-tification of *Pseudomonas aeruginosa* and *Aspergillus fumigatus* mono- and co-cultures based on volatile biomarker combinations, J. Breath Res. 10 (1) (2016) 016002.

[11] C. Roullier, S. Bertrand, E. Blanchet, M. Peigné, T. Robiou du Pont, Y. Guitton, Y.F. Pouchus, O. Grovel, Time dependency of chemodiversity and biosynthetic pathways: an LC-MS metabolomic study of marine-sourced penicillium, Mar. Drugs 14 (5) (2016) 103.

[12] S. Bertrand, O. Schumpp, N. Bohni, A. Bujard, A. Azzollini, M. Monod, K. Gindro, J.L. Wolfender, Detection of metabolite induction in fungal co-cultures on solid media by high-throughput differential ultra-high pressure liquid chro-matography–time-of-flight mass spectrometry fingerprinting, J. Chromatogr. A 1292 (0) (2013) 219–228.

[13] S. Bertrand, O. Schumpp, N. Bohni, M. Monod, K. Gindro, J.L. Wolfender, De

[14] novo production of metabolites by fungal co-culture of trichophyton rubrum and bionectria ochroleuca, J. Nat. Prod. 76 (6) (2013) 1157–1165.

[14] R.A. Janzen, S.B. Rood, J.F. Dormaar, W.B. McGill, Azospirillum brasilense produces gibberellin in pure culture on chemically-defined medium and in co-culture on straw, Soil Biol. Biochem. 24 (10) (1992) 1061–1064.

[15] D.C. Oh, C.A. Kauffman, P.R. Jensen, W. Fenical, Induced production of emer-icellamides A and B from the marine-derived fungus emericella sp. in competing Co-culture, J. Nat. Prod. 70 (4) (2007) 515–520.

[16] E. Saccenti, H.C. Hoefsloot, A.K. Smilde, J.A. Westerhuis, M.M. Hendriks, Re-flections on univariate and multivariate analysis of metabolomics data, Metabolomics 10 (3) (2014) 361–374.

[17] A. Combès, I. Ndoye, C. Bance, J. Bruzaud, C. Djediat, J. Dupont, B. Nay, S. Prado, Chemical communication between the endophytic fungus Paraconiothyrium variabile and the phytopathogen Fusarium oxysporum, PLos One 7 (10) (2012) e47313.

[18] A.W. Boots, A. Smolinska, J.J.B.N. van Berkel, R.R.R. Fijten, E.E. Stobberingh, M.L.L. Boumans, E.J. Moonen, E.F.M. Wouters, J.W. Dallinga, F.J. Van Schooten, Identification of microorganisms based on headspace analysis of volatile organic compounds by gas chromatography–mass spectrometry, J. Breath Res. 8 (2) (2014) 027106.

[19] D.K. Derewacz, B.C. Covington, J.A. McLean, B.O. Bachmann, Mapping micro-bial response metabolomes for induced natural product discovery, ACS Chem. Biol. 10 (9) (2015) 1998–2006.

[20] J.R. de Haan, R. Wehrens, S. Bauerschmidt, E. Piek, R.C. van Schaik, L.M.C. Buydens, Interpretation of ANOVA models for microarray data using PCA, Bioinformatics 23 (2) (2007) 184–190.

[21] P.D.B. Harrington, N.E. Vieira, J. Espinoza, J.K. Nien, R. Romero, A.L. Yergey, Analysis of variance–principal component analysis: a soft tool for proteomic discovery, Anal. Chim. Acta 544 (1–2) (2005) 118–127.

[22] J. Engel, L. Blanchet, B. Bloemen, L.P. van den Heuvel, U.H.F. Engelke, R.A. Wevers, L.M.C. Buydens, Regularized MANOVA (rMANOVA) in untargeted metabolomics, Anal. Chim. Acta 899 (2015) 1–12.

[23] J.J. Jansen, H.C. Hoefsloot, J. van der Greef, M.E. Timmerman, J.A. Westerhuis, A.K. Smilde, ASCA: analysis of multivariate data obtained from an experi-mental design, J. Chemom. 19 (9) (2005) 469–481.

[24] A.K. Smilde, J.J. Jansen, H.C. Hoefsloot, R.J.A. Lamers, J. Van Der Greef, M.E. Timmerman, ANOVA-simultaneous component analysis (ASCA): a new tool for analyzing designed metabolomics data, Bioinformatics 21 (13) (2005) 3043–3048.

[25] P.J. Van den Brink, C.J. Ter Braak, Principal response curves: analysis of time-dependent multivariate responses of biological community to stress, Environ. Toxicol. Chem. 18 (2) (1999) 138–148.

[26] H.C. Keun, T.M. Ebbels, M.E. Bollard, O. Beckonert, H. Antti, E. Holmes, J.C. Lindon, J.K. Nicholson, Geometric trajectory analysis of metabolic re-sponses to toxicity can define treatment specific profiles, Chem. Res. Toxicol. 17 (5) (2004) 579–587.

[27] A.K. Smilde, M.E. Timmerman, M.M. Hendriks, J.J. Jansen, H.C. Hoefsloot, Generic framework for high-dimensional fixed-effects ANOVA, Brief. Bio-inform. 13 (5) (2012) 524–535.

[28] J.J. Jansen, R. Bro, H.C. Hoefsloot, F.W. van den Berg, J.A. Westerhuis, A.K. Smilde, PARAFASCA: ASCA combined with PARAFAC for the analysis of metabolic fingerprinting data, J. Chemom. 22 (2) (2008) 114–121.

[29] F. Marini, D. de Beer, E. Joubert, B. Walczak, Analysis of variance of designed chromatographic data sets: the analysis of variance-target projection approach, J. Chromatogr. A 1405 (2015) 94–102.

[30] S. Wold, K. Esbensen, P. Geladi, Principal component analysis, Chemom. Intelligent Laboratory Syst. 2 (1–3) (1987) 37–52.

[31] R. Bro, A.K. Smilde, Principal component analysis, Anal. Methods 6 (9) (2014) 2812–2831.

[32] T.W. Chippendale, F.J. Gilchrist, P. Španěl, A. Alcock, W. Lenney, D. Smith, Quantification by SIFT-MS of volatile compounds emitted by in vitro cultures of S. aureus, S. pneumoniae and H. influenzae isolated from patients with respiratory diseases, Anal. Methods 6 (8) (2014) 2460–2472.

[33] J.J. Jansen, H.C. Hoefsloot, J. Van Der Greef, M.E. Timmerman, A.K. Smilde, Multilevel component analysis of time-resolved metabolic fingerprinting data, Anal. Chim. Acta 530 (2) (2005) 173–183.

[34] J.J. Jansen, N.M. van Dam, H.C. Hoefsloot, A.K. Smilde, Crossfit analysis: a novel method to characterize the dynamics of induced plant responses, BMC Bio-inform. 10 (1) (2009) 1–15.

[35] Y. Xu, S.J. Fowler, A. Bayat, R. Goodacre, Chemometrics models for overcoming high between subject variability: applications in clinical metabolic profiling studies, Metabolomics 10 (3) (2013) 375–385.

[36] G. Zwanenburg, H.C. Hoefsloot, J.A. Westerhuis, J.J. Jansen, A.K. Smilde, ANOVA–principal component analysis and ANOVA–simultaneous component analysis: a comparison, J. Chemom. 25 (10) (2011) 561–567.

[37] G. Luciano, T. Næs, Interpreting sensory data by combining principal component analysis and analysis of variance, Food Qual. Prefer. 20 (3) (2009) 167–175.

[38] T. Pluskal, S. Castillo, A. Villar-Briones, M. Orešič, MZmine 2: modular framework for processing, visualizing, and analyzing mass spectrometry-based molecular profile data, BMC Bioinform. 11 (1) (2010) 395.

[39] J.L. Wolfender, G. Marti, A. Thomas, S. Bertrand, Current approaches and challenges for the metabolite profiling of complex natural extracts, J. Chromatogr. A 1382 (2015) 136–164.

[40] L.W. Sumner, A. Amberg, D. Barrett, M.H. Beale, R. Beger, C.A. Daykin,

T.W.M. Fan, O. Fiehn, R. Goodacre, J.L. Griffin, T. Hankemeier, Proposed minimum reporting standards for chemical analysis, Metabolomics 3 (3) (2007) 211—221.

[41] T. Kind, O. Fiehn, Seven golden rules for heuristic filtering of molecular formulas obtained by accurate mass spectrometry, BMC Bioinform. 8 (1) (2007) 1.

[42] Chapman and Hall, Dictionary of Natural Products on DVD (23:1), vol. 23, CRC Press, Taylor & Francis Group, 2014, 1, http://dnp.chemnetbase.com/.

[43] A.H. Neerincx, J. Mandon, J. van Ingen, D.D. Arslanov, J.W. Mouton, F.J. Harren, P.J. Merkus, S.M. Cristescu, Real-time monitoring of hydrogen cyanide (HCN) and ammonia (NH3) emitted by *Pseudomonas aeruginosa*, J. Breath Res. 9 (2) (2015) 027102.

[44] D.J. Vis, J.A. Westerhuis, A.K. Smilde, J. Van Der Greef, Statistical validation of megavariate effects in ASCA, BMC Bioinform. 8 (1) (2007) 1—8.

[45] B. Briard, C. Heddergott, J.-P. Latgé, Volatile compounds emitted by *Pseudomonas aeruginosa* stimulate growth of the fungal pathogen *Aspergillus fumigatus*, mBio 7 (2) (2016) e00219, 16.