



# **A MALDI–MS Profiling to Address Honey Bee Health Status under Bacterial Challenge through Computational Modeling**

Karim Arafah, Sébastien Nicolas Voisin, Victor Masson, Cédric Alaux, Yves Le Conte, Michel Bocquet, Philippe Bulet

## **► To cite this version:**

Karim Arafah, Sébastien Nicolas Voisin, Victor Masson, Cédric Alaux, Yves Le Conte, et al.. A MALDI–MS Profiling to Address Honey Bee Health Status under Bacterial Challenge through Computational Modeling. *Proteomics*, 2019, 19 (23), pp.1900268. <10.1002/pmic.201900268>. <hal-02994660>

**HAL Id: hal-02994660**

**<https://hal.science/hal-02994660v1>**

Submitted on 18 Nov 2020

**HAL** is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.



HAL Authorization

A MALDI-MS profiling to address honey bee health status under bacterial challenge through  
computational modelling

Karim ARAFAH<sup>1</sup>, Sébastien Nicolas VOISIN<sup>1</sup>, Victor MASSON<sup>1,2</sup>, Cédric ALAUX<sup>3</sup>, Yves Le  
CONTE<sup>3</sup>, Michel BOCQUET<sup>4</sup>, Philippe BULET<sup>1,2\*</sup>

<sup>1</sup>Plateforme BioPark Archamps, Forum 1, 260 Avenue Marie Curie  
Archamps Technopole, F-74160 Saint Julien en Genevois, France

<sup>2</sup>CR UGA, Institute for Advanced Biosciences, Inserm U1209, CNRS UMR5309, Team Analytic  
immunology of Chronic diseases, Site Santé, Allée des Alpes, F-38000 Grenoble, France

<sup>3</sup>INRA PACA, UR 406 Abeilles et Environnement, 228 route de l'aérodrome, Site Agroparc, Domaine  
Saint-Paul, 84914 Avignon Cedex 9, France

<sup>4</sup>Apimedia, BP22 -Pringy- 74371 Annecy cedex, France

\*Corresponding Author, [philippe.bulet@univ-grenoble-alpes.fr](mailto:philippe.bulet@univ-grenoble-alpes.fr); secondary email:  
[philippe.bulet@biopark-archamps.org](mailto:philippe.bulet@biopark-archamps.org), direct line [+33\(0\)4 50 43 25 21](tel:+33(0)450432521)

- 8166 words (including references and figure legends)
- 8010 words (excluding first page, abstract, significance of the study, acknowledgments,  
declaration of interest)
- Abstract: 200 words
- Significance of the study: 184 words

## Abstract:

Honey bees play a critical role in the maintenance of plant biodiversity and sustainability of food webs. In the last few decades, bees have been subjected to biotic and abiotic threats causing various colony disorders. Therefore, monitoring solutions to help beekeepers to improve bee health are necessary. MALDI mass spectrometry profiling has emerged within this decade as powerful to identify in routine micro-organisms and is currently used in real-time clinical diagnosis. We developed MALDI BeeTyping to monitor significant hemolymph molecular changes in honey bees upon infection with a series of entomopathogenic Gram-positive and -negative bacteria. A *Serratia marcescens* strain isolated from one naturally infected honey bee collected from the field was also considered. We individually recorded a series of hemolymph molecular mass fingerprints and built, to our knowledge, the first computational model harboring a predictive score of 97.92% and made of nine molecular signatures that discriminate and classify the honey bees' systemic response to the bacteria. Hence, we challenged our model by classifying a training set of hemolymphs and obtained an overall recognition of 91.93%. Through this work, we aimed at introducing a novel, time and cost saving high-throughput strategy that addresses honey bee health and on an individual scale.

## Keywords:

Bee health, MALDI profiling, Bacterial infection, *Apis mellifera*, Computational model

## Statement of significance of the study:

Domesticated and wild bees worldwide represent the most active and valuable pollinators that ensure plant biodiversity and the success of many crops. These pollinators and others are exposed to deleterious pathogens and environmental stressors acting alone or in combination. Despite efforts to better understand how these threats affect honey bee (*Apis mellifera*) health status, solutions are still crucially needed to help beekeepers, scientists and other stakeholders in obtaining either a prognosis, an early diagnosis or a diagnosis of the health status of the apiaries. In this study, we describe a new method to investigate honey bee health through a simple "blood test". The fingerprinting of endogenous peptides and proteins contents of honey bees' hemolymph gives a molecular signature reflecting health status in presence of infections by entomopathogenic bacteria. Based on a reproducible protocol involving the establishment of well characterized models of infection, we automated the identification of infected bees with a predictive score of 97.92%. The proposed molecular mass fingerprinting of hemolymph, named MALDI BeeTyping as a MALDI profiling strategy, could be a new, time and cost saving method to monitor honey bee's health.

## Introduction

Over several decades, an abnormal mortality of honey bees and other pollinators (bumblebees, solitary bees) has been observed in all industrialized countries<sup>[1-4]</sup>. This phenomenon has been particularly

recorded in honey bees (*Apis mellifera*)<sup>[5]</sup>. The global loss of honey bee colonies has detrimental consequences for plant biodiversity, bee products, and negative economic and societal effects<sup>[6]</sup>. As a result, many scientific studies have been carried out to understand the mechanisms underlying phenomena such as colony weakening or collapse and colony mortality observed in most of the countries practicing intensive agriculture. Many reports concluded that biotic and abiotic factors are suspected to be involved in this phenomenon, either alone or in combination<sup>[2, 5, 7-10]</sup>. Potential causes are exposure to (i) environmental and in-hive chemicals<sup>[11, 12]</sup>, (ii) agricultural practices<sup>[13, 14]</sup>, (iii) infection by micro-organisms and predation by parasites<sup>[15-17]</sup> and (iv) nutritional factors<sup>[18-20]</sup>, among others, which lead to the transition from a health status qualified as normal to a health decline that would contribute to the colony collapse<sup>[7]</sup>. The expression of this pathological state may notably be linked to a decrease in the immune capacities of the bee and/or the colony<sup>[21-25]</sup>. The complex underlying mechanisms of stressors (biotic and abiotic) that impact honey bee health status remain still partially understood and solutions capable of rendering a prognosis, an early diagnosis or a diagnosis, need to be developed. Even if visual inspection of beehives, polymerase chain reaction and sensor-based analyses are available for surveillance of pathogen loads, prediction of the likely impact on the colony remains an issue not satisfactorily addressed<sup>[26-29]</sup>. Taking advantages of analytical improvements over the last 20 years, MALDI-MS has become increasingly popular to identify biological samples through mass fingerprinting of molecular changes that occur in fluid, cells or insects' appendices or simply as a robust microbiological technique in clinical diagnostics<sup>[30]</sup>. This approach may also play an essential role in the quest for innovative solutions in monitoring bee health.

Molecular mass fingerprinting (MFP) or signatures of biological matrices by MALDI-MS profiling is indeed a thriving approach, enabling the rapid detection of peptide/protein that can provide comparative information regarding immunity in *Drosophila*<sup>[31-33]</sup>, insects' appendices<sup>[34-36]</sup> and even from complex biological matrices. MALDI-MS based peptidomics/proteomics typing has found practical applications in fingerprinting honeys to determine specific molecular signatures that classify honeys according to their geographical origins<sup>[37]</sup> or to establish a robust proteomics mass fingerprint of sterile sperm chicken for diagnosis purpose<sup>[38]</sup>.

Relying on previously published studies<sup>[33, 39-41]</sup>, we first developed and validated an experimental model of challenged honey bees with Gram-positive and -negative bacteria (using notably a *Serratia marcescens* strain we have isolated from honey bees). Then, we assessed the usefulness of MALDI profiling to fingerprint the peptides/proteins within the molecular mass range 2-20 kDa from honey bee hemolymph. Finally, we generated a computational model that scored nine markers as a discriminant signature. These markers were successfully used to discriminate hemolymph molecular changes that occurred following bacterial infections. Through the present work, we are introducing this newly developed approach of MALDI profiling, referred as MALDI BeeTyping, to follow the health status of *Apis mellifera* under bacterial infections. While clinical application of biotyping is based on the identification of proteins derived from the micro-organisms themselves (for more about the biotyping

approach, see<sup>[42-44]</sup>), the BeeTyping approach we describe in this paper does not aim neither at identifying proteins from bacteria present in the hemolymph nor at isolating and characterizing bacteria from this fluid but rather at identifying the subsequent molecular changes they induced in the honey bee hemolymph. In addition, we determined the performance of this computational model using a training set of challenged bees. Through this work, we introduce BeeTyping as an effective method for monitoring peptide/protein changes within honey bees' hemolymph following bacterial challenges as a blood.

## Material & Methods

The MALDI profiling strategy relies on a workflow divided into four major steps (see Supporting Information 1) and described in this section.

### Biological models

#### Bacterial strains

To generate biological models of infection, we used the Gram-negative strains *Pectobacterium carotovorum* subsp. *carotovorum* 15 (formerly *Erwinia carotovora* 15 CFBP2141, generous gift from Bruno Lemaitre, EPFL Switzerland), *Serratia entomophila* (Institut Pasteur, CIP102919) and the Gram-positive *Micrococcus luteus* (ATCC 4698). We added in our infection model, a *Serratia marcescens* strain (SmBIOP160412, from our laboratory collection) isolated within the haemocoel from a naturally infected *Apis mellifera* collected in the field. This strain was identified using MALDI Biotyper® Compass (v4.1 from Bruker Daltonik GmbH) and according to the Bruker manufacturer protocol. Bacteria were cultured in Luria Bertani (LB) medium overnight at 32°C.

#### Experimental infection of the honey bees

Experimental infections were performed on newly-emerged honey bee workers (less than 12h old). To design the computational analyses, a training set of spectra was built using non experimentally infected (unpicked control) bees and bees infected with either *Pectobacterium carotovorum* subsp. *carotovorum* 15, *M. luteus*, *S. entomophila* or the isolated *Serratia marcescens* SmBIOP160412 strain. Infections were performed by pricking honey bees individually in the anterior lateral thorax (spiracle) using a fine needle (Fine Science Tools, Germany) dipped into a freshly concentrated culture pellet of live bacteria. All honey bees (experimentally infected and controls) were placed for 24h at room temperature in dedicated small cages and fed *ad libitum* with sugar syrup (Invertbee from SARL Isnard, France) containing fructose (36%), dextrose (30%), saccharose (31%), maltose (1.5%) and other sugars (1.5%). Hemolymph was collected from the dorsal side of the abdomen, using pulled glass capillaries (Sutter Instrument

132 Corp, Novato, California). The collected hemolymph was immediately transferred into a  
133 chilled LoBind Protein microtube (Eppendorf, Germany) pre-coated with phenylthiourea  
134 and phenylmethanesulfonyl fluoride (both from Sigma Aldrich, France) to prevent  
135 melanization and proteolysis, respectively. The hemolymph samples were stored at -20°C  
136 until use.

137

## **Molecular mass fingerprints by MALDI MS**

### **Data acquisition**

Each individual hemolymph sample was analyzed with the Bruker AutoFlex™ III. The molecular mass fingerprints (MFP) were acquired following the Bruker Biotyper® recommendations (matrix, method of sample deposition and detection) with minor adjustments. Briefly, the hemolymph samples were 10-fold diluted in acidified water (0.1% trifluoroacetic acid, Sigma Aldrich, France) and 0.5µL of a given sample was mixed with 0.5µL of 4-HCCA (Sigma Aldrich, France) on a MALDI MTP 384 polished ground steel plate (Bruker Daltonik). Following co-crystallization of the hemolymph spots with the matrix droplet, MALDI MS spectra were recorded in a linear positive mode and in an automatic data acquisition using FlexControl 4.0 software (Bruker Daltonik). The following instrument settings were used: 1.5kV of electric potential difference, dynamic range of detection of 600 to 18,000 Da, 69% of laser power, a global attenuator offset of 46% with 200Hz laser frequency, and 2,000 accumulated laser shots per hemolymph spectrum with a raster of random walk set to 50. The linear detector gain was setup at 1.82kV with a suppression mass gate up to  $m/z$  600 to prevent detector saturation by clusters of the 4-HCCA matrix. The pseudo-molecular ions desorbed from the hemolymph were accelerated under 1.5kV. An external calibration of the mass spectrometer was performed using a standard mixture of peptides and proteins (Peptide Standard Calibration II and Protein Standard Calibration I, Bruker Daltonik) covering the dynamic range of analysis.

### **Data post-processing and statistical analyses**

The MALDI-MS datasets were imported into the ClinProTools™ 2.2 Software (Bruker Daltonik) for post-processing and statistical analyses. All of the recorded spectra were processed with a baseline subtraction and spectral smoothing followed by an internal recalibration step with exclusion of null and/or “non-recalibratable” spectra. The total averaged spectra were calculated based on a signal over noise ratio equal to 5 for peak-picking and area calculations. The irrelevant spectra that did not pass the required signal intensity and resolution were excluded from any integration into the MALDI-MS computational model designed to match the biological models of honey bee infections. A post-processing step involving spectral normalization of all calculated peak area was performed with ClinProTools™ software prior to statistical analysis (95% confidence interval, standard deviation and Principal Component Analysis-PCA).

### *Hierarchical Clustering, heat maps and Receiver Operating Characteristic curves*

The total number of spectra used to design the computational models were normalized and subjected to PCA and unsupervised hierarchical clustering analysis to measure distances between spectra. This analysis was used to determine Euclidean distances (based on PCA results with a reduced dimension limited to 70% and 95% of the total explained variance). The molecular correlation between four antimicrobial peptides (AMPs) known from the honey bee [Apidaecin 1A at  $m/z$  2,107 (Uniprot entry A0A088AIG0), Hymenoptaecin at  $m/z$  10,270 (Uniprot entry Q10416), Abaecin at  $m/z$  3,878 (Uniprot entry P15450) and Defensin 1A at  $m/z$  5,519 (Uniprot entry P17722)]; and mass fingerprints (MFP) of the three biological models of infection were calculated and represented with a heat map. The ROC analyses were built using the ClinProTools™ program and the heat maps, using the OMICs add-on module provided by the XLSTAT program (interquartile threshold value of 0.25).

#### Computational-based algorithm & machine learning model

In the quest of delivering a barcode model capable of discriminating infected from control honey bees, a training set of spectra was established by fingerprinting the corresponding hemolymph samples using MALDI MS. Series of individual spectra were recorded from 22 controls, 23 honey bees challenged with *M. luteus*, 20 with *P. carotovorum* subsp. *carotovorum* 15, 15 with *S. entomophila* and 15 with *S. marcescens* SmBIOP160412. Data clustering (optimal spectral separation combined with the determination of a fixed number of peaks within the training set) was performed using the Genetic Algorithm (GA) with the ClinProTools™ software. The GA parameters were as follow: a maximum of 10 peaks harboring the greatest weight was selected and included in the model. A number of 50 generations (iterative algorithm searching) was chosen to achieve this maximum of peaks. The k-nearest neighbor parameter, which is a key parameter of artificial intelligence used in supervised machine learning, was set at 3.

#### External validation of the barcode model and classification of unknown spectra

In order to assess the capability of the GA classifier to recognize the infected bees from the control group, a new set of hemolymph MS spectra, never processed in the classifier model, was used to perform an external validation. This experimental set of honey bees included the three biological models; 26 controls, 37 honey bees infected with *P. carotovorum* subsp. *carotovorum* 15 and 10 with *M. luteus*. By submitting those hemolymph spectra to the classifier resulted in counting the correctly classified spectra, and also the mismatched and the invalid ones. In order to assess the performance of our classifier model, accuracy, sensitivity, specificity, informedness, specific-positive and negative likelihood ratios, false discovery rate (q-value) and false positive rate (p-value) were calculated. The accuracy, which informs



207 on how efficient the model is, was calculated according to Wang *et al.*<sup>[45]</sup>. Sensitivity scores real positive  
208 cases that are correctly predicted positive by the model and the specificity scores the opposite i.e. the  
209 real negative cases that are correctly predicted negative. Informedness scores the probability that a  
210 prediction (e.g. result of a machine-learning model to classify one condition against the others) is  
211 informed regarding to the tested condition *versus* odds. Informedness helps to make diagnosis decision.  
212 Sensitivity, specificity and informedness were determined as previously described<sup>[46]</sup>. The specific-  
213 positive and -negative likelihood ratios (abbreviated +LR and -LR) classically used in diagnostic  
214 testing with multiple classes informs on how likely the results from the classifier model will  
215 match the condition. +LR gives the change in the odds of satisfying the condition (fitting to the  
216 biological models), given a positive test result and -LR, the change in the odds of satisfying the  
217 condition when the test comes negative. +LR ranges from zero to infinity. With +LR values  
218 between zero and one, there is a weak probability that the test matches the condition. If the ratio  
219 equals to one, then the test lacks diagnostic value and if the ratio is  $>1$ , then the test increases  
220 the probability to match correctly with the condition. Regarding -LR, the closer to zero the  
221 value is, the more informative the test is<sup>[47]</sup>.

## Results & Discussion

### MALDI-MS profiling as a new approach to discriminate bacterial systemic infection through endogenous proteins from honey bees fingerprinting

In order to generate relevant biological models of honey bee infections, we used the bacteria *Micrococcus luteus*, *Pectobacterium carotovorum* subsp. *carotovorum* and *Serratia entomophila*. *M. luteus* is a Gram-positive bacterial strain frequently used when monitoring insect immunity<sup>[48]</sup> and has been shown to colonize bee hives and gastrointestinal tracts of honey bees<sup>[49]</sup> while *P. carotovorum* subsp. *carotovorum*<sup>[50]</sup> and *S. entomophila* (Institut Pasteur, CIP102919) are two bacterial strains that trigger a systemic immune response in insects. We also performed an additional model of infection using a *Serratia marcescens* strain (SmBIOP160412, Lab. collection) isolated from a naturally infected honey bee collected in the field and which is known to be a widespread pathogen of adult honey bees<sup>[51]</sup>, and a virulent opportunist, taking advantage of disturbed microbiota to develop in honey bee guts after exposure to the pesticide glyphosate<sup>[52]</sup>.

To certify the constructed biological models of infection, the four bacterial strains were classified by MALDI MS biotyping (Data not shown). A set of 64 MALDI MS spectra was recorded from individual hemolymph samples. These spectra were obtained from 22 control honey bees and from 23 and 19 honey bees individually infected with the Gram-positive *M. luteus* or the Gram-negative *P. carotovorum* subsp. *carotovorum*, respectively. An averaged spectrum, containing 110 MALDI MS ion peaks (MFP, Supporting Information 2), was built for each of the three biological models (Figure 1A). Statistical analysis based on Principal Component Analysis (PCA) and performed on these MFPs clearly segregated the three biological models (Figure 1B). As shown by the PCA plot score, the individual spectra were clustered in accordance with their corresponding models and were segregated based on their mass fingerprints. The unsupervised hierarchical clustering of hemolymph samples, classified almost all of the individual MFPs with respect to their corresponding biological models (Figure 1C). Out of 64 normalized spectra used to build the clustering dendrogram, four and three recorded mismatched spectra were observed, corresponding to the lowest (70%) and highest (95%) limits of explained variance, respectively. The mismatched spectra were further identified within the representation of the PCA plot score of the hemolymph samples (Figure 1B, arrows). At the limit of explained variance of 70%, the mismatched four spectra included one spectrum from the control condition and one from the *M. luteus* infection model, both classified under the *P. carotovorum* subsp. *carotovorum* model, and two of this latter model, mismatched to the control model (see asterisks in Fig. 2C). Regarding the three mismatched spectra observed at 95% of the explained variance, one was from the *M. luteus* model and classified under the *P. carotovorum* subsp. *carotovorum* model and two, from the *P. carotovorum* subsp. *carotovorum* model, classified under the control model (see asterisks in Fig. 2C).

In order to assess the relationship between the MALDI-MS MFPs of the biological models and the honey bee's immune status, we correlated these MFPs with each of the four antimicrobial peptides (AMP) defined from *Apis mellifera*<sup>[53]</sup>: Apidaecin 1A (41) at  $m/z$  2,107, Hymenoptaecin<sup>[54]</sup> at  $m/z$  10,270, Abaecin<sup>[55]</sup> at  $m/z$  3,878 and Defensin 1A<sup>[39]</sup> at  $m/z$  5,519 (Figure 2). As shown, per-peak fingerprint correlations with the antimicrobial peptides (AMPs) were scored based on the molecular ion peak area and represented as heat maps through a colored scale intensity ranging from low (minimum score of -1, in red) to high correlation (maximum score of 1, in green). Reported in relation to the MFPs, four clades (A, B, C and D) described the positive and negative correlations of the MFPs with each of the four AMPs (see Supporting Information 2) and segregated the three biological models (non-experimentally infected as a control condition, *P. carotovorum* subsp. *carotovorum*. model and *M. luteus* model, see Figure 2).

Regarding the control condition, the four AMPs were found to be positively correlated with the molecular ion markers of the hemolymph MFPs of clades A, C and D and negatively correlated with markers of clade B. In the *P. carotovorum* subsp. *carotovorum* model, the same four AMPs were positively correlated with the MFPs of clades B and C and negatively with clade A and D markers, except for Hymenoptaecin, which exhibited positive and negative correlations with clade D. In the *M. luteus* model of infection, each of the AMPs was predominantly positively correlated with the molecular clades A and B, and negatively with clades C and D. These correlations show complementary molecular signatures in the three experimental models. Discrete dynamic molecular patterns are modulated and correlated to the immune status of the bees, allowing us to discriminate infected from non-infected bees and the type of infection.

### **Machine-learning as the first reported computational model to recognize and classify experimentally infected honey bees based on hemolymph MFPs**

Because the proteomic mass spectra of hemolymph samples reflect the immune status of the honey bees, our next goal was to predict honey bee health status based on the bee MALDI-MS MFPs. For that purpose, we decided to build a molecular model based on the MFPs of hemolymph samples, by using a machine-learning algorithm, the Genetic Algorithm (GA). The GA classifier generated a set of discriminating peaks that recognized and classified hemolymph according to the biological model (honey bees challenged with *P. carotovorum carotovorum* or *M. luteus* and non-experimentally infected honey bees). These discriminating peaks form a barcode model and define the strength of this model through its recognition capability. The performance of the classifier barcode model was evaluated through internal cross validation by iterative reclassification of a set of spectra equal to half of the total number of spectra included in the model. For each of the ten iterations performed, a new set of spectra was chosen randomly through an automated internal process.

In an initial approach, we restricted our experimental infection to *M. luteus* as the Gram-positive strain and to *P. carotovorum* subsp. *carotovorum*. While this is, to our knowledge, the first time such a computational model has been applied to the classification of bacterial-infected honey bees, machine-learning algorithms have been used previously in other biological subjects. For example, MALDI MS has been successfully used to build a proteomic mass spectra database of different honeys and their MFPs to identify their geographical origin<sup>[37]</sup>. As another example of application, an experimental model of male chicken fertility was designed to perform on-cell direct proteomic fingerprinting by MALDI profiling and demonstrated the capability of the GA classifier to build a predictive model to classify chicken sperm fertility<sup>[38]</sup>.

In the present study, using GA, based on the individual hemolymph spectra of a cohort of 22 controls and 23 honey bees challenged with *M. luteus* or 20 honey bees challenged with *P. carotovorum* subsp. *carotovorum*, we identified a set of nine best  $m/z$  molecular ions based on their capability to discriminate the three biological models from each other (Figure 3). Further tests of recognition capability and cross validation of the GA model were assessed by using the MFPs from the same sample cohort. Considering the standard deviation and the 95% confidence interval of these nine molecular ions, weight indexes were calculated to rank the nine molecular signatures from the most discriminant molecular ion ( $m/z$  3,348.17, weight of 6.24) to the least discriminant one ( $m/z$  5,603.01, weight of 1.97). Moreover, we rated the accuracy of the GA classifier model following two distinct data processing. On the one hand, the classifier calculated the recognition capability by matching the MALDI MS spectra described above against their respective biological models. Therefore, we were able to re-assign hemolymph spectra derived from the control and the *M. luteus* biological models (score of recognition 100%) and for the *P. carotovorum* subsp. *carotovorum* model (score of 93.75%). Overall, performance recognition of the classifier reached 97.92%. On the other hand, internal cross validation scores were calculated for each biological model. To perform this cross validation, the same individual hemolymph spectra from each biological model were randomized and reassessed for successful matching in a batch mode of analysis by the classifier using solely the set of the nine molecular ion markers. The cross validations of the classifier were at 91.51%, 94.40% and 89.87% for the control, *M. luteus* and *P. carotovorum* subsp. *carotovorum* biological models, respectively, giving an overall validation of 91.93% (see Figure 3).

A new set made of 26, 10 and 35 MALDI MS spectra of hemolymph from control, *M. luteus* and *P. carotovorum* subsp. *carotovorum* biological models respectively was submitted for the to the GA classifier and classified individually (Table 1 and Supporting Information 3). Among the 26 control spectra, 16 were correctly classified, three were classified in *P. carotovorum* subsp. *carotovorum* and three in *M. luteus* models. Four spectra were found as invalid spectra because of the recalibration step. This result was caused by weaker intensities of the molecular fingerprints causing the ion mass recalibration to fail. Regarding the *M. luteus* infectious model, 10 spectra were subjected to the classifier. Seven were correctly classified, two were considered as control and one as belonging to the *P. carotovorum* subsp. *carotovorum* model. No spectrum was deemed invalid. Considering the *P.*

*carotovorum* subsp. *carotovorum* biological model of infection, from the 37 spectra, 13 were correctly classified, one matched to the control, three to the *M. luteus* biological models and 20 to the invalid spectra category. These 20 spectra were qualified as invalid due to noisy mass spectra (intensities of the nine peaks not sufficient to pass the classification) or to a failure in properly calibrating the mass spectra. Given these results, we calculated the performance of the classifier for each of the biological model (see details in Supporting Information 4). The GA algorithm achieved 80 % to 90 % accuracy discriminating thus the three biological models. The sensitivity (true positive) and the specificity (true negative) of the GA classifier model were calculated for the three biological models. The model scored at least 70% of sensitivity and at least 84 % of specificity. As detailed in the Supporting Information 4, the highest sensitivity was observed for the *P. carotovorum* subsp. *carotovorum* model (76.47 %) and the highest specificity for the control model (95.16 %). Based on the sensitivity and the specificity, we calculated the informedness indexes and the positive-negative stratum-specific likelihood ratio (abbreviated +LR, -LR) which inform about how predictive the classifier model is and its performance as a diagnostic tool respectively. As reported in the Supporting Information 4, the three biological models scored indexes within the range [-1 ;1] with -1 as incorrect model predictions, 1 as maximum of correct predictions). The calculated informedness indexes for the control, the *P. carotovorum* subsp. *carotovorum* scoring 0.68 and 0.64 respectively and for the *M. luteus* (0.55) demonstrated the model was a good predictor. Regarding +LR and -LR, both parameters were calculated. The +LR, which required scores over 1 to be significant were found equal to 15.02; 4.55 and 6.12 for the control, the *M. luteus* and the *P. carotovorum* subsp. *carotovorum* models respectively (Supporting Information 4). This result demonstrated a good probability that our GA model classified positively the spectra against the biological models. The -LR, which required scores as close as possible to 0 to be significant were found equal to 0.28; 0.35 and 0.27 for the control, the *M. luteus* and the *P. carotovorum* subsp. *carotovorum* models respectively (Supporting Information 4). This result demonstrated the weak probability to misclassify the cohort of hemolymph spectra through the GA classifier. In addition, we determined the false discovery rate (q-value) and the false positive rate (p-value) for each of the three biological models. The lowest q-value was of 0.158 and concerned the control model while *P. carotovorum* subsp. *carotovorum* and *M. luteus* models harbored highest values (0.235 and 0.461 respectively) denoting a better capability of the classifier to classify unknown spectra within the control model followed by the *P. carotovorum* subsp. *carotovorum* and *M. luteus* models respectively. Regarding the p-values, the control model harbored 0.0484 while *P. carotovorum* subsp. *carotovorum* and *M. luteus* models scored 0.125 and 0.154. Based on this statistical parameter, the classifier shares the same conclusion as obtained with the q-values regarding the algorithm's performance to classify properly the spectra. To summarize, the computational model significantly discriminated control from infected honey bees and *M. luteus* from *P. carotovorum* subsp. *carotovorum* infection as well on the basis of the hemolymph molecular signatures.

## **MALDI-MS profiling of hemolymph as an effective molecular method to discriminate honey bees infected with different *Serratia* species**

We assessed the performance of MALDI profiling to discriminate honey bee infections at the species level, in particular between two *Serratia* species: *S. marcescens* (SmBIOP160412, Lab. collection) isolated from a naturally infected honey bee and identified by MALDI Biotyper® Compass (data not shown), and a reference strain of *S. entomophila*. We found that the MFPs of the hemolymph samples collected from honey bees infected by *S. entomophila* and *S. marcescens* presented significant molecular differences (Figure 4A).

Moreover, by using the two best discriminant molecular ions ( $m/z$  12,752.8 and  $m/z$  7,186.95), revealed by the PCA analysis, we could differentiate the hemolymph spectra of bees infected either with *S. entomophila* or *S. marcescens* SmBIOP160412. In contrast, the two weaker discriminant markers  $m/z$  1,996.14 and  $m/z$  6,113.68 failed to discriminate the two types of spectra resulting from the two *Serratia* species (Figure 4B).

To further evaluate and rank the measured ion markers within the MFPs based on their capability to discriminate the two *Serratia* species, we performed a receiver operating curve (ROC) analysis to highlight eight  $m/z$  ion markers (12,752.8; 7,186.95; 7,688.27; 10,269.8 (Hymenoptaecin); 5,057.69; 5,160.28) with AUC scores between 0.8 and 1 in sensitivity (Figure 4C). The Apidaecin, Abaecin and Defensin were also checked for their capability to discriminate the two *Serratia* species (Supporting Information 5). Based on the ROC test, Abaecin and Defensin were also capable to discriminate, to some extent, the two *Serratia* species (see Supporting Information 5) (AUC of 0.739 and 0.639, respectively). However, like the two markers  $m/z$  1,996.14 and 6,613.68, Apidaecin (AUC=0.520) revealed to be a poor discriminant (Figure 4D). Hence, based on these results, it seems possible to discriminate infection by *S. marcescens* from *S. entomophila* in honey bees using computational modelling. To our knowledge, this is the first report on the feasibility of using MALDI MS as an MFP-based method capable of discriminating hemolymph molecular response to systemic infections induced by two different bacterial species of the same genus. *S. marcescens* is known to be a commensal bacterium present in low abundance in the gut of honey bees. By studying the pathogenicity of different strains of *S. marcescens* through two routes of *in vivo* exposure (oral and direct injection into the hemolymph),

Raymann *et al.* (46) found that expression of the four honey bee AMPs Abaecin, Defensin, Hymenoptaecin and Apidaecin did not differ between infected and non-experimentally infected control honey bees. These results support the idea that markers other than AMPs need to be identified and monitored to efficiently discriminate bacterial infections in honey bee hemolymph. As we demonstrated, the correlation of the molecular fingerprints

and the AMPs in hemolymph allowed us to discriminate the three different models. In order to determine how specific were the nine markers of the classifier to the control, *M. luteus* and *P. carotovorum* subsp. *carotovorum* biological models, we tested the classification of the hemolymph mass fingerprints of honey bees infected using the two *Serratia* strains. We submitted 13 MALDI mass fingerprints of hemolymph from infected bees with *S. entomophila* to the classifier. One was classified as control, seven as fitting with the *P. carotovorum* subsp. *carotovorum* model, and one fitting with the *M. luteus* model. The classifier excluded four hemolymphs' fingerprints because of noisy spectra signal or invalid mass recalibration. We also submitted 19 MALDI mass fingerprint of hemolymph from infected bees with *S. marcescens* to the classifier, which identified two as control, five as fitting with the *P. carotovorum* subsp. *carotovorum* model, and four fitting with the *M. luteus* model. The classifier excluded eight hemolymphs' fingerprints for the same reasons as above. Interestingly, the majority of the classified spectra from both, the *S. entomophila* and *S. marcescens* models matched with the *P. carotovorum* subsp. *carotovorum* model. It is particularly interesting as these three bacteria are Gram-negative. Nevertheless, some spectra matched against the control and the *M. luteus* models. Taking altogether, these results suggest that the MALDI profiling approach generates specific molecular barcodes/signatures defined accordingly to biological models. As we developed this approach for the monitoring of honey bee health, we define this approach of molecular signature by MALSI MS as MALDI BeeTyping.

## Concluding remarks

Along with most relevant, technically feasible and primary observation-based health status indicators highlighted by EFSA's HEALTHY-B, MALDI-MS BeeTyping, a method derived from the biotyping approach used routinely in clinical microbiology, analyze the downstream responses to stressors through the matured effector molecules circulating in the hemolymph. These effectors include the products of selected immune genes (i.e. genes coding for Apidaecin, Defensin, Hymenoptaecin, and Abaecin) and other molecular markers of stress that we intend to further characterize through an on-going proteomic study. Our approach of MFPs by MALDI-MS BeeTyping, is a cutting-edge analytic method that may complement and address some limitations issued of the HEALTHY-B toolbox by establishing robust, effective, sensitive and a comprehensive technology for profiling and deciphering, at the individual level, the honeybee health parameters including its immunity stage with regards to bacterial stressors. Moreover, MALDI BeeTyping has several advantages over other molecular biology techniques and visual observations, such as (i) the use of a drop of hemolymph allowing to keep the rest

of the body for complementary molecular measurements such as PCR, (ii) a very simple and fast sample preparation, (iii) a short time for data acquisition and processing, (iv) low consumable costs, and (v) a user friendly workflow that can be standardized and automated for cost-effective high throughput use. We believe that future developments of MALDI BeeTyping could improve monitoring of honey bee health upon exposure to other biotic or abiotic stressors, the quality control and the origin traceability of apiary products based on molecular markers fingerprinting. Based on specific proteomics signatures, MALDI BeeTyping could bring out a novel analytical tool for early diagnosis of honey bees parasited with *Nosema* species, *Varroa destructor* and infected or not with deformed wing virus or acute bee paralysis virus.



447 **Acknowledgements**

448 We would like to thank Dr. Bruno Lemaitre from the Ecole Polytechnique Fédérale de Lausanne –  
449 (EPFL, Switzerland) and the Institut Pasteur de Paris for their generous gifts (*Micrococcus luteus* strain  
450 and *Serratia entomophila* respectively). We are also thankful to Paul-Arthur Sol and Erwan Weber for  
451 their contribution during their Master 2 training. Financial support was provided by FEAGA  
452 (FranceAgriMer), N°14-04R and from the Association Plateforme BioPark of Archamps from its  
453 Research & Development budget.

454

455 **Conflict of interest**

456 The authors have declared no conflict of interest.

## References

- [1] J. C. Biesmeijer, S. P. Roberts, M. Reemer, R. Ohlemuller, M. Edwards, T. Peeters, A. P. Schaffers, S. G. Potts, R. Kleukers, C. D. Thomas, J. Settele, W. E. Kunin, *Science* 2006, 313, 351.
- [2] S. G. Potts, J. C. Biesmeijer, C. Kremen, P. Neumann, O. Schweiger, W. E. Kunin, *Trends in ecology & evolution* 2010, 25, 345.
- [3] A. J. Vanbergen, *Front Ecol Environ* 2013, 11, 251.
- [4] I. Koh, E. V. Lonsdorf, N. M. Williams, C. Brittain, C., R. Isaacs, J. Gibbs, T. H. Ricketts, *Proc Natl Acad Sci U S A.* 2016, 113, 140.
- [5] D. Goulson, E. Nicholls, C. Botias, E. L. Rotheray, *Science* 2015, 347, 982.
- [6] T. D. Breeze, B. E. Vaissiere, R. Bommarco, T. Petanidou, N. Seraphides, L. Kozák, J. Scheper, J. C. Biesmeijer, D. Kleijn, S. Gyldenkaerne, M. Moretti, A. Holzschuh, I. Steffan-Dewenter, J. C. Stout, M. Pärtel, M. Zobel, S. G. Potts, *PloS One* 2014, 9, e82996.
- [7] K. M. Smith, E. H. Loh, M. K. Rostal, C. M. Zambrana-Torrel, L. Mendiola, P. Daszak, *EcoHealth* 2013, 10, 434.
- [8] B. Dainat, D. Vanengelsdorp, P. Neumann, *Environ Microbiol Rep*, 2012, 4, 123.
- [9] P. Neumann, N. L. Carreck, *J Apic R* 2010, 49, 1.
- [10] N. Steinhauer, K. Kulhanek, K. Antunez, H. Human, P. Chantawannakul, M. P. Chauzat, D. Vanengelsdorp, *Curr Opin Insect Sci* 2018, 26, 142.
- [11] C. A. Mullin, M. Frazier, J. L. Frazier, S. Ashcraft, R. Simonds, D. Vanengelsdorp D, J. S. Pettis, *PloS One* 2010, 5, e9754.
- [12] J. Y. Wu, M. D. Smart, C. M. Anelli, W. S. Sheppard, *J Invertebr Pathol* 2012, 109, 326.
- [13] A. Clermont, M. Eickermann, F. Kraus, L. Hoffmann, M. Beyer, *Sci Total Environ* 2015, 532, 1.
- [14] M. Smart, J. Pettis, N. Rice, Z. Browning, M. Spivak, *PloS One* 2016, 11, e0152685.
- [15] M. Beyer, J. Junk, M. Eickermann, A. Clermont, F. Kraus, C. Georges, A. Reichart, L. Hoffmann, *Res Vet Sci* 2018, 118, 52.
- [16] B. E. Traver, H. K. Feazel-Orr, K. M. Catalfamo, C. C. Brewster, R. D. Fell, *J Econ Entomol* 2018, 111, 517.
- [17] S. D. Desai, R. W. Currie, *PloS One* 2016, 11, e0159615.
- [18] G. Di Pasquale, M. Salignon, Y. Le Conte, L. P. Belzunces, A. Decourtye, A. Kretzschmar, S. Suchail, J. L. Brunet, C. Alaux, *PloS One* 2013, 8, e72016.
- [19] C. Alaux, C. Dantec, H. Parrinello, Y. Le Conte, *BMC genomics* 2011, 12, 496.
- [20] A. G. Dolezal, A. L. Toth, *Curr Opin Insect Sci* 2018, 26, 114.
- [21] C. Sinpoo, R. J. Paxton, T. Disayathanoowat, S. Krongdang, P. Chantawannakul, *J Insect Physiol.* 2018, 105, 1.
- [22] F. Nazzi, D. Annoscia, E. Caprio, G. Di Prisco, F. Pennacchio, *Entomologia* 2014, 2, 80.
- [23] M. L. Flenniken, R. Andino, *PloS One* 2013, 8, e77263.
- [24] Y. Zhang, X. Liu, W. Zhang, R. Han, *J Insect Physiol.* 2010, 56, 1207.
- [25] G. Di Prisco, V. Cavaliere, D. Annoscia, P. Varricchio, E. Caprio, F. Nazzi, G. Gargiulo, F. Pennacchio, *Proc Natl Acad Sci U S A.* 2013, 110, 18466.
- [26] EFSA, *EFSA Journal* 2016, 14, 1.
- [27] W. Meikle, N. Holst, *Apidologie* 2015, 46, 10.
- [28] S. Gil-Lebrero, F. J. Quiles-Latorre, M. Ortiz-Lopez, V. Sanchez-Ruiz, V. Gámiz-López, J. J. Luna-Rodríguez, *Sensors* 2016, 17, pii: E55.
- [29] M. A. Becher, G. Twiston-Davies, T. D. Penny, D. Goulson, E. L. Rotheray, J. L. Osborne, *J Appl Ecol* 2018, 55, 2790.
- [30] J. Rychert, *J Infectiology* 2019, 2, 1.
- [31] L. Sabatier, E. Jouanguy, C. Dostert, D. Zachary, J. L. Dimarcq, P. Bulet, J. L. Imler, *Eur J Biochem* 2003, 270, 3398.
- [32] N. Boulanger, L. Ehret-Sabatier, R. Brun, D. Zachary, P. Bulet, J. L. Imler, *Insect Biochem Mol Biol* 2001, 31, 129.

507 [33] S. Uttenweiler-Joseph, M. Moniatte, M. Lagueux, A. Van Dorsselaer, J. A. Hoffmann, P. Bulet, Proc  
508 Natl Acad Sci U S A 1998, 95, 11342.

509 [34] A. Vega-Rua, N. Pages, A. Fontaine, C. Nuccio, L. Hery, D. Goindin, J. Gustave, L. Almeras, Parasit  
510 Vectors 2018, 11, 574.

511 [35] P. H. Boyer, N. Boulanger, A. Nebbak, E. Collin, B. Jaulhac, L. Almeras, PLoS One 2017, 12,  
512 e0185430.

513 [36] D. S. Loginov, Y. F. Loginova, F. Dycka, K. Bottinger, P. Vechtova, J. Sterba, Parasit Vectors 2019,  
514 12, 212.

515 [37] J. Wang, M. M. Kliks, W. Qu, S. Jun, G. Shi, Q. X. Li, J Agric Food Chem 2009, 57, 10081.

516 [38] L. Soler, V. Labas, A. Thelie, I. Grasseau, A. P. Teixeira-Gomes, E. Blesbois, Mol Cell Proteomics  
517 2016, 15, 1998.

518 [39] K. Casteels-Josson, W. Zhang, T. Capaci, P. Casteels, P. Tempst, J Biol Chem 1994, 269, 28569.

519 [40] P. Casteels, C. Ampe, F. Jacobs, M. Vaeck, P. Tempst, EMBO J 1989, 8, 2387.

520 [41] K. Ishii, H. Hamamoto, K. Sekimizu, PLoS One 2014, 9, e89917.

521 [42] J. S. Scott, S. A. Sterling, H. To, S. R. Seals, A. E. Jones, Infect Dis (Lond) 2016, 48, 530.

522 [43] M. Kostrzewa, Expert Rev Proteomics 2018, 15, 193.

523 [44] K. S. Jang, Y. H. Kim, J Microbiol 2018, 56, 209.

524 [45] H. Y. Wang, F. Lien, T. P. Liu, C. H. Chen, C. J. Chen, J. J. Lu, PeerJ 2018, 6, e5784.

525 [46] D. M. W. Powers, J Mach Learn Technol 2011, 2, 37.

526 [47] S. McGee, J Gen Intern Med 2002, 17, 647.

527 [48] C. Hetru, P. Bulet, Methods Mol Biol 1997, 78, 35.

528 [49] P. Pomastowski, M. Zloch, A. Rodzik, M. Ligor, PLoS One 2019, 14, e0217078.

529 [50] C. Acosta Muniz, D. Jaillard, B. Lemaitre, F. Boccard, Cell Microbiol 2007, 9, 106.

530 [51] K. Raymann, K. L. Coon, Z. Shaffer, S. Salisbury, N. A. Moran, mBio 2018, 9.

531 [52] E. V. S. Motta, K. Raymann, N. A. Moran, Proc Natl Acad Sci U S A 2018, 115, 10305.

532 [53] J. Danihlík, K. Aronstein, M. Petřivalský, J Apic R 2015, 54, 123.

533 [54] P. Casteels, C. Ampe, F. Jacobs, P. Tempst, J Biol Chem 1993, 268, 7044.

534 [55] P. Casteels, C. Ampe, L. Riviere, J. Van Damme, C. Elicone, M. Fleming, F. Jacobs, P. Tempst, Eur J  
535 Biochem 1990, 187, 381.

## Figure legends

**Figure 1:** Differential PCA-based statistical analyses and hierarchical clustering of individual hemolymph samples from the biological models.

Total averaged spectra were fingerprinted by MALDI MS from the infected and control individuals (**A**). The individual spectra were subjected to PCA analysis, which discriminated the hemolymph molecular mass fingerprints of *Micrococcus luteus* (*M. l.* in blue), *Pectobacterium carotovorum* subsp. *carotovorum* 15 (*P. c. c.* in green) and control (red) groups (**B**); arrows mark the mismatched outliers. An unsupervised hierarchical clustering based on the PCA results classified the individual spectra according to the lowest (70% left panel **C**) and highest (95%, right panel **C**) limits of explained variances.

**Figure 2:** Heat-map of four antimicrobial peptides (AMPs) from *Apis mellifera*, Apidaecin 1A, Abaecin, Defensin 1A and Hymenoptaecin correlating with the MALDI MS fingerprints (102 *m/z*) of the hemolymph samples.

Per-peak MALDI MS correlation in standard mode between the AMP molecular ions of Apidaecin 1A (*m/z* 2,107), Abaecin (*m/z* 3,878), Defensin 1A (*m/z* 5,519) and Hymenoptaecin (*m/z* 10,270), and the MALDI MS fingerprints of the biological model following an experimental infection with either *Micrococcus luteus* (*M. l.*) in blue, or *Pectobacterium carotovorum* subs. *carotovorum* (*P. c. c.*) in green, and the control experiment (non-experimentally infected bees, in red). Each rectangle in the heat map dendrogram represents the abundance level (scale from -1 to +1 from the lowest in red to the highest in green, respectively) of the area of each AMP cross-related with each molecular ion from the fingerprint.

**Figure 3:** Genetic Algorithm-based classifier used to discriminate non-experimentally infected (Control, red) bees from experimentally infected ones with *Micrococcus luteus* (*M. l.*, in blue) or *Pectobacterium carotovorum* subsp. *carotovorum* (*P. c. c.*, in green).

Nine molecular ion peaks determined by the computational model and ranked according to their weight indexes were found as the best discriminative features of the hemolymph samples on the basis of statistical criteria (Standard deviation determined for each curve representing the molecular ions of the model and the box plots showing the first and third interquartile range with line denoting the median and whiskers encompassing 95% of the individuals). The spectral gel view represents the intensity of each of the discriminative peaks represented by their weight index and the *m/z* values (Da) found within the individual spectra of the biological models.

**Figure 4:** Differential PCA-based analysis of hemolymph fingerprints following infection with two *Serratia* strains, *Serratia marcescens* (*S. m.*) isolated from a naturally infected *Apis mellifera* and *S. entomophila* (*S. e.*) and statistical relevance of predictive markers.

The differential fingerprinting and PCA analysis discriminated the non-experimentally infected bees (control in black) and the experimentally infected groups (*S. e.* in red or *S. m.* in green) with n=13 bees per group (A). The detected peaks in the differential analysis with the highest and lowest discriminant scores were assessed statistically by measuring the standard deviation and the 95% confidence interval. (B). The most interesting peaks classified through the Receiver Operating Characteristics (ROC) curves are shown using the Area Under Curve (AUC) calculation (C). The biological model used as the positive class was the experimental *S. m.* infection and the sensibility (True Positives) and specificity (False Positives) parameters were determined for all calculated peaks. Eight markers defined by their *m/z* values (Da) were found as the best predictive markers (AUC>0.8) for discriminating honey bee infections (*S. e.* or *S. m.*). Conversely, the two least discriminant peaks had an irrelevant AUC (~0.5) with ROC curves fitting the non-discriminant line of the statistical test (D).

**Table 1:** External validation of the genetic algorithm classifier model using a new set of hemolymph spectra.

Fifty-seven hemolymph samples were collected individually from the biological models *M. l.*, *P. c. c.* and control prior to being fingerprinted. The spectra were submitted to the GA-based computational model in order to assess classifier performance.

**Supporting Information 1:** MALDI profiling workflow for machine learning data-driven analysis of honey bee infections.

The methodological approach relied on four main steps addressing major tasks.

Step 1: Sampling of unchallenged bees (controls) and experimental infection obtained by pricking honey bees with live strains of *Pectobacterium carotovorum* subsp. *carotovorum* 15 (*P. c. c.*) and *Micrococcus luteus* (*M. l.*). Step 2: Individual hemolymph collections followed by MALDI-MS molecular mass fingerprinting, and strain identification by MALDI biotyping. Step 3: Multi-stage processing of MALDI MS fingerprints including recalibration, peak picking, normalization and statistical calculation of individual MS spectra through Principal Component Analysis (PCA) for revealing differential molecular patterns across infection groups. Step 4: Genetic algorithm-based computational model for recognition and classification of honey bee infection using PCA discriminant analysis. Barcodes were built following the molecular fingerprints that discriminate control bees from bees infected either with *M. l.* or with *P. c. c.*

**Supporting Information 2:** Peak-to-peak correlation scores of the four antimicrobial peptides (Abaecin, Apidaecin, Defensin and Hymenoptaecin) with the mass fingerprint of hemolymph across the three biological models (non-experimentally infected/control, *M. l.* for *Micrococcus luteus* infection and *P. c. c.* for *Pectobacterium carotovorum* subsp. *carotovorum* 15 infection).

**Supporting Information 3:** Results for the external validation of the genetic algorithm-based classifier.

609 **Supporting Information 4:** Assessment of the Genetic Algorithm classifier performance.  
610 Based on the result of the external validation, the performance of the GA-based classifier model was  
611 assessed for each of the biological models by calculating the accuracy, the sensitivity, the specificity,  
612 the specific-positive and -negative likelihood ratios (all five expressed as percentage), informedness, p-  
613 value and q-value.

614 **Supporting Information 5:** Assessment of ROC curves of Apidaecin, Abaecin and Defensin to  
615 discriminate *S. marcescens*- from *S. entomophila*-infected honey bees.