

RESEARCH ARTICLE

Bees can be trained to identify SARS-CoV-2 infected samples

Evangelos Kontos^{1,2,*}, Aria Samimi^{1,*}, Renate W. Hakze-van der Honing³, Jan Priem³, Aurore Avarguès-Weber⁴, Alexander Haverkamp², Marcel Dicke², Jose L. Gonzales³ and Wim H. M. van der Poel^{3,‡}

ABSTRACT

The COVID-19 pandemic has illustrated the need for the development of fast and reliable testing methods for novel, zoonotic, viral diseases in both humans and animals. Pathologies lead to detectable changes in the volatile organic compound (VOC) profile of animals, which can be monitored, thus allowing the development of a rapid VOC-based test. In the current study, we successfully trained honeybees (*Apis mellifera*) to identify SARS-CoV-2 infected minks (*Neovison vison*) thanks to Pavlovian conditioning protocols. The bees can be quickly conditioned to respond specifically to infected mink's odours and could therefore be part of a wider SARS-CoV-2 diagnostic system. We tested two different training protocols to evaluate their performance in terms of learning rate, accuracy and memory retention. We designed a non-invasive rapid test in which multiple bees are tested in parallel on the same samples. This provided reliable results regarding a subject's health status. Using the data from the training experiments, we simulated a diagnostic evaluation trial to predict the potential efficacy of our diagnostic test, which yielded a diagnostic sensitivity of 92% and specificity of 86%. We suggest that a honeybee-based diagnostics can offer a reliable and rapid test that provides a readily available, low-input addition to the currently available testing methods. A honeybee-based diagnostic test might be particularly relevant for remote and developing communities that lack the resources and infrastructure required for mainstream testing methods.

KEY WORDS: Honeybees, Olfaction, SARS-CoV2, Covid-19, Conditioning, Detection

INTRODUCTION

Infections and other pathologies lead to physiological changes in the bodies of animals (Trabue et al., 2010) and humans (Buljubasic and Buchbauer, 2015; Sethi et al., 2013; Shirasu and Touhara, 2011; Probert et al., 2009). Consequently, the emitted volatile organic compounds (VOCs) differ between healthy and infected individuals (Wilson, 2018; Fitzgerald et al., 2017; Olsson et al., 2014; Trabue et al., 2010; Probert et al., 2009). VOCs constitute an odour fingerprint depending on age, sex, diet, genetic background, and metabolic conditions, thus making this odour fingerprint unique for

every individual (Buljubasic and Buchbauer, 2015). Analysing that fingerprint can provide relevant information about the state of the individual's health. VOC analysis has been consequently used for disease diagnostics, mostly in the form of breath and faeces analysis in both humans and animals (Wilson, 2018; Fitzgerald et al., 2017; Olsson et al., 2014; Trabue et al., 2010; Probert et al., 2009).

The current COVID-19 pandemic has clearly shown the need for both the rapid development of diagnostic tests and the rapid delivery of reliable results (Control, 2020). Fast and reliable diagnostic tests are required to effectively implement control measures such as quarantine of infected people or animals (Wells et al., 2021) There is a global need for reliable and rapid testing, which has led to the development of very reliable PCR tests and rapid SARS-CoV-2 tests such as the RNA RT-LAMP (Fowler et al., 2021) and antigen tests (Krüttgen et al., 2021). However, in developing countries and remote areas, such methods may not be easily available. Dogs have been successfully trained to discriminate between SARS-CoV-2-infected and non-infected individuals with a diagnostic sensitivity ranging from 65% to 82.6% and specificity of 89% and 96.4% respectively (Eskandari et al., 2021; Jendry et al., 2020). Similar to dogs, some insects have keen olfactory capabilities. For example, fruit flies (*Drosophila melanogaster*) can detect cancer in humans (Strauch et al., 2014), while honeybees (*Apis mellifera*) have exhibited the ability to detect some human diseases, such as tuberculosis (Suckling and Sagar, 2011). Honeybees can, therefore, be a potential alternative to dogs for the detection of COVID-19 with the benefit of being readily available and having low costs of maintenance.

Pavlovian conditioning was first applied to dogs (Pavlov, 1927) and later to honeybees (Takeda, 1961). Bees possess the reflex to extend their proboscis when detecting a sugar solution, the proboscis extension reflex (PER), and they can be conditioned to exhibit a PER when exposed to specific odours. Takeda's (1961) classical conditioning pairs a conditioning stimulus (CS), such as an odour, with an unconditioned stimulus (US), the food reward, which in most cases is a sugar water solution (Matsumoto et al., 2012). After such training the bees exhibit PER when exposed to the CS, without the presence of sugar water.

Previous studies have shown that animals can detect differences between VOCs emitted by healthy or SARS-CoV-2 infected individual animals or humans (Eskandari et al., 2021; Jendry et al., 2020). The objective of this study was to assess the potential of training bees for the detection of SARS-CoV2-infected animal samples. We assessed two different training methods and show that bees can be effectively trained to detect differences in odours between samples collected from SARS-CoV2 infected and uninfected minks (*Neovison vison*), highlighting the potential of a honeybee-based diagnostic test for the detection of diseases.

RESULTS

Protocol 1

We analysed the bees' learning curve during conditioning by fitting a generalized linear mixed model (GLMM). A significant interaction

¹InsectSense, Plus Ultra-II Building, Bronland, 10, 6708 WH, Wageningen, The Netherlands. ²Laboratory of Entomology, P.O. Box 16, 6700 AA, Wageningen, The Netherlands. ³Wageningen Bioveterinary Research, P.O. Box 65, 8200 AB, Lelystad, The Netherlands. ⁴Centre de Recherches sur la Cognition Animale (CRCA), Centre de Biologie Intégrative (CBI), Université de Toulouse; CNRS, UPS, 118 Route de Narbonne, 31062 Toulouse, France.

*These authors contributed equally to this work

‡Author for correspondence (WimvanderPoel1@wur.nl)

© A.H., 0000-0003-3512-9659; M.D., 0000-0001-8565-8896; J.L.G., 0000-0002-7889-5380; W.H.M.v., 0000-0002-7498-8002

This is an Open Access article distributed under the terms of the Creative Commons Attribution License (<https://creativecommons.org/licenses/by/4.0>), which permits unrestricted use, distribution and reproduction in any medium provided that the original work is properly attributed.

Table 1. Logistic regression mixed model results analysing the bees' learning curves

Predictors	Protocol 1				Protocol 2				
	Log-odds	SE	Z	P	Log-odds	SE	Z	P	
Intercept [Infected]	-1.61	0.36	-4.5	<0.001	-1.58	0.22	-7.15	<0.001	
Sample [Healthy]	-1.06	0.49	-2.13	0.033	0	0.26	0	0.997	
Conditioning round	0.31	0.06	4.86	<0.001	0.36	0.04	10.22	<0.001	
Sample [Healthy]* Conditioning round	-0.38	0.11	-3.37	0.001	-0.26	0.05	-5.48	<0.001	
Random Effects									
Residual SD	1.81				1.81				
Bee id SD	1.47				1.09				
Number of bees	56				92				
Observations	784				1582				
Conditional R ²	0.551				0.388				

SD: standard deviation.

[log-odds=-0.38, standard error (SE)=0.11, Z=-3.37, $P<0.001$] between treatment and conditioning round was observed, which suggests a significant increase in the bee's ability to discriminate between infected and negative samples with increasing conditioning

rounds (Table 1). By the end of the conditioning phase (round 7), 37 bees out of the total 56 bees (66.1%) expressed PER towards the infected sample (old-infected) and 4 out of 56 (7.1%) towards the healthy sample (old-healthy) (Fig. 1A).

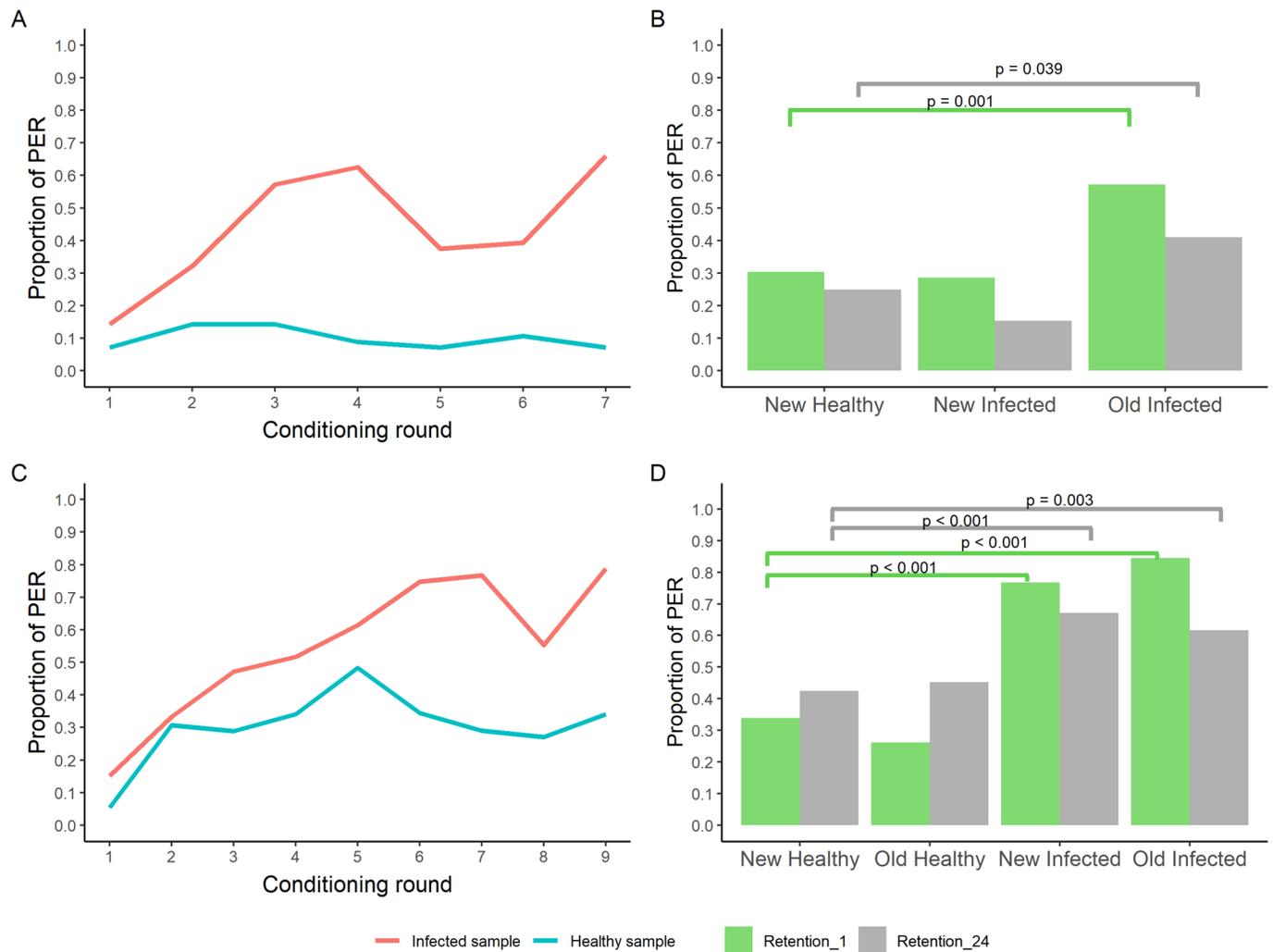


Fig. 1. Learning and memory retention of the bees. (A,B) show the learning curve ($n=56$ bees) and memory retention ($n=56$ bees) of bees subjected to protocol 1. (C,D) show the learning curve ($n=92$ bees) and memory retention ($n=56$ bees) of bees subjected to protocol 2. In panels A and C (learning curves), the Y-axis shows the proportion of bees expressing PER towards infected (red) and healthy (blue) samples in each conditioning round while the X-axis indicates the conditioning round. In panels C and D (memory retention), the Y-axes show the proportion of bees expressing PER and the X-axes show the different types of samples that the bees were exposed to 1 h (green columns) and 24 h (grey columns) after the conditioning training ended. Segments and corresponding P values indicate comparisons where significant. The sample type, new-healthy was used as reference for statistical comparison.

We tested the bees' memory retention 1 h after the conditioning phase. Every bee was exposed to three odours (samples): old-infected (sample used for conditioning), new-infected and new-healthy. During exposure, the old-healthy sample was always present as a background odour. Using the GLMM analysis confirmed that most of the bees were able to discriminate (log-odds=1.8, SE=0.56, Z=3.24, $P=0.001$) the old-infected sample [32 out of 56 (57.1%) reacted to old-infected] from the new-healthy sample [17 out of 56 (30.4%) reacted to new-healthy]. However, bees were not able to discriminate (log-odds=-0.13, SE=0.52, Z=-0.26, $P=0.796$) the new-infected sample [16 out of 56 (28.6%)] from the new-healthy sample (Fig. 1B).

Another retention test took place 24 h after the end of training, using the same odours we had used for the 1 h retention test. When analysing the reaction of the bees 24 h after conditioning, out of 56 bees, 14 (25%) reacted to the new-healthy sample, 23 reacted to the old-infected (41.1%) sample and 8 out of 52 (15.4%) reacted to the new-infected sample. Bees were only able to significantly discriminate between the old-infected (log-odds=1.06, SE=0.51, Z=2.07, $P=0.038$) and the new-healthy samples (Fig. 1B).

Protocol 2

Similar to protocol 1, we analysed the bees' learning curve during conditioning using a GLMM. A significant interaction (log-odds=-0.26, SE=0.05, Z=-5.48, $P<0.001$) between treatment and conditioning round was observed, which suggests a significant increase in the bee's discrimination ability between the positive and negative samples with the conditioning rounds (Table 1). By the end of conditioning (round 9) 67 out of 85 (78.8%) bees expressed PER towards infected samples and 23 out of 85 (27.1%) towards healthy samples (Fig. 1C).

We tested the bees' memory retention 1 h after the end of the conditioning. Overall, 71 bees out of 84 (84.5%) reacted to the old-infected sample. Bees reacted 129 times out of 168 trials (77.8%) to the new-infected samples while 22 bees out of 84 (26.2%) reacted to the old-healthy samples. Bees reacted 57 times out of 168 trials (33.9%) to new-healthy samples. The GLMM analysis (using new-healthy as reference for comparison) confirmed that bees were able to significantly discriminate (express PER) between the new-healthy sample and both the old (log-odds=3.02, SE=0.42, Z=7.27, $P<0.001$) and new-infected (log-odds=2.41, SE=0.31, Z=7.79, $P<0.001$) samples. No differences were observed between the bees' reaction towards the old- (log-odds=-0.47, SE=0.33, Z=-1.40, $P=0.161$) and new-healthy samples (Fig. 1D).

Another retention test was executed 24 h after the end of training, using the same odours we had used for the 1 h retention test. Overall, 45 bees out of 73 (61.6%) reacted to the old-infected sample and 98 times out of 146 trials (67.1%) to the new-infected samples, while 33 bees out of 73 (45.2%) reacted to the old-healthy samples and 62 times out of 146 trials (42.5%) to new-healthy samples. The GLMM analysis showed that bees were able to significantly discriminate between the new-healthy and both the old-infected (log-odds=1.01, SE=0.34, Z=3.02, $P=0.003$) and new-infected (log-odds=1.31, SE=0.28, Z=4.59, $P<0.001$) samples. The bees' reaction to the old and new-healthy samples did not differ significantly (log-odds=0.14, SE=0.33, Z=0.44, $P=0.658$) (Fig. 1D).

Comparison between the sequences (punishment first, reward first), did not result in a sequence that yielded a significantly better results during 1 h. Whereas during the 24 h retention only one comparison was significantly different (data not shown). This

indicated that the sequence with which the samples were provided to the bees did not significantly affect their training outcome.

Bees as a diagnostic tool

Association between Ct values and the bees' retention ability

We compared the proportion of bees showing PER depending on the Ct values of the infected samples. After being trained on a sample with a Ct=21, a total of 42 bees were exposed to samples with three different Ct values (Ct: 21, 27 or 30): 35 (83.3%) bees showed PER for the sample with a Ct=21, 40 (95.2%) showed the samples with Ct=27 and 31 (73.8%). The samples used for training were those with a Ct=21, hence we used this group as reference for comparison. No significant differences were observed between either the Ct=21 group and the Ct=27 group ($\chi^2=0.63$, d.f.=1, $P=0.16$, $P_{adjusted}=0.475$) or Ct=21 and Ct=30 ($\chi^2=0.63$, d.f.=1, $P=0.42$, $P_{adjusted}=1$) groups, indicating that bees trained with samples having a high virus concentration (a low Ct value) are still able to recognize samples with lower virus concentrations (high Ct value).

Predicted performance when using bees as a diagnostic tool

The distribution of diagnostic results when testing healthy and infected samples in a simulated scenario in which a group of 10 trained bees would be used to test a sample, is shown in Fig. 2A. The receiver operating characteristics (ROC) analysis on the simulated data resulted in an estimated area under the curve (AUC) (Fig. 2B) equal to 0.96 (95% confidence interval, CI): 0.95–0.98), which indicates that using groups of trained bees could be a diagnostic tool with significant discrimination accuracy (AUC>0.5, $P<0.001$). Using a response of six or more (out of 10) bees showing PER per test to classify a sample as positive would maximize the diagnostic performance of this tool. The resulting potential sensitivity (true positive rate), which is the probability that the test will correctly classify a truly infected sample as positive, would be 0.92 (95% CI: 0.89–0.95) and the potential specificity (true negative rate), which is the probability that the test will correctly classify a healthy sample as negative, 0.86 (95% CI: 0.82–0.90) (Fig. 2).

DISCUSSION

The objective of this study was to investigate whether bees can be trained to detect SARS-CoV-2 infected samples. Our data show that the differences in odour between SARS-CoV-2 infected samples and uninfected samples can be recognised by honeybees. The bees discriminated between samples taken from healthy and SARS-CoV2 infected individuals. Although the bees' discrimination ability decreases between 1 h and 24 h post conditioning, we observed that they were nevertheless still able to significantly discriminate between new infected and healthy samples 1 day after conditioning. Moreover, their ability to recognise a positive sample was not compromised by the samples' viral load (expressed in Ct values), since bees recognised samples with higher Ct values (lower viral load) equally well as they did with samples with low Ct values used for conditioning. By performing simulations of the potential clinical application of the bees as a diagnostic tool, we predict that bees could be effective for diagnostics with a predicted sensitivity around 92% and specificity around 86%.

With our protocol 1, we followed a similar procedure as Sutherland et al. (2010). In our study, more bees learned to recognize the rewarded odour (current study: 66.1%, Sutherland et al.: 30–40%). Sutherland et al. (2010) reported that one hour after training, 20% of the trained bees were no longer able to discriminate, which is similar to the reduction in percentage we

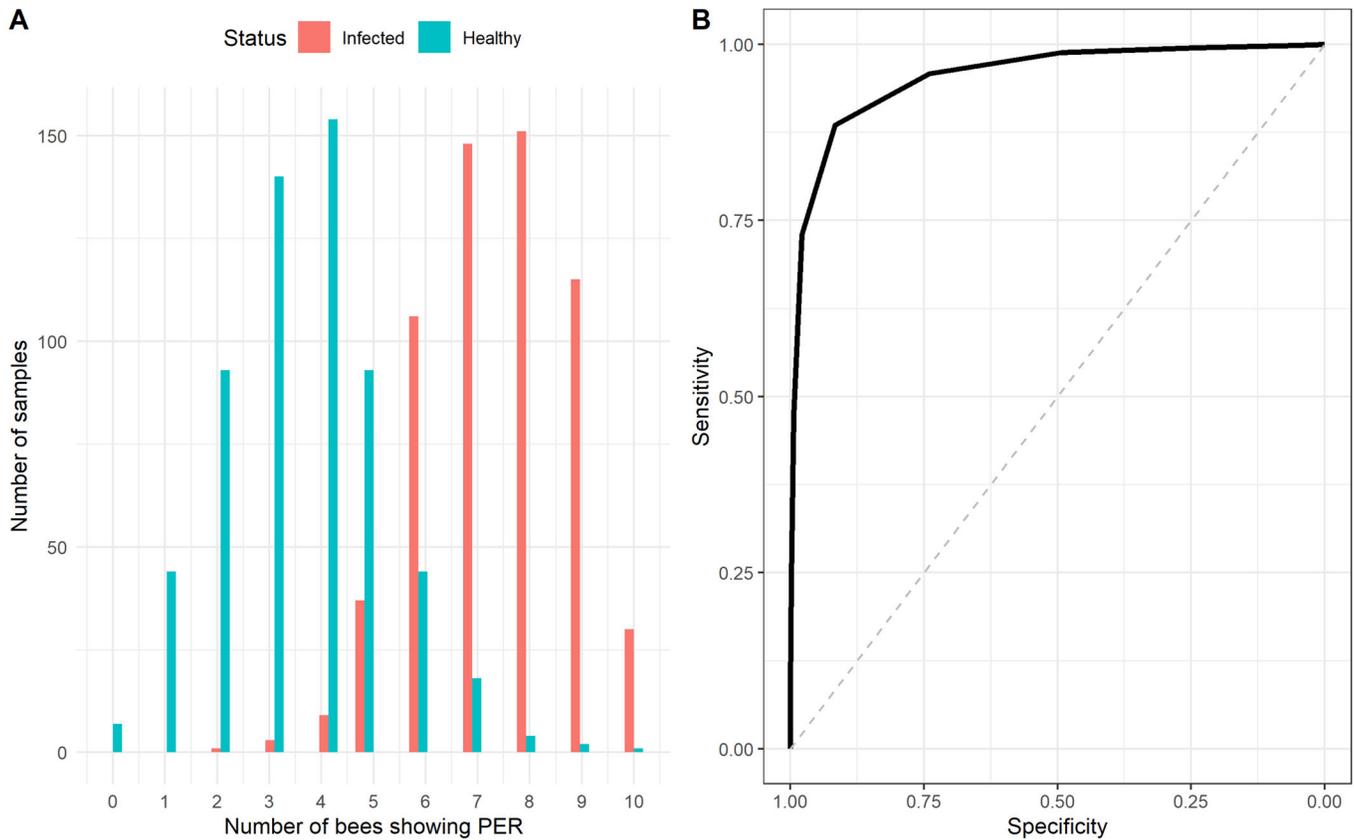


Fig. 2. Simulated diagnostic potential of trained bees. (A) Distribution of simulated diagnostic results where a group of 10 bees is used as diagnostic tool per sample. The X-axis indicates the number of bees (out of 10) per test showing PER. (B) ROC of the predicted diagnostic performance.

observed. Sutherland et al. (2010) did not perform any test with novel infected and healthy samples, and we did not find literature where a similar approach to ours was taken. In protocol 2, we followed a similar procedure as Wright et al. (2010). The bees in the present experiment learned slightly less well than in their experiment (current study: 78.8%, Wright et al.: >85%). Assuming that this difference is significant, we could speculate that this is a result of samples being more difficult to discriminate either by being more perceptually similar or less concentrated. Wright et al. (2010) did not test the 1 h memory of the bees, but only their memory after 10 min. Their results were similar to ours at 1 h (current study: 83.5%, Wright et al.: 80%).

Protocol 1 was shorter than protocol 2, making it a faster way to condition bees, while it also required no aversive unconditioned stimulus (US) and fewer samples during training. However, protocol 1 did not result in the bees being able to discriminate between the novel-infected and novel-healthy samples. That indicates that they were not able to generalize between infected samples or to associate VOCs that commonly occur in infected samples with a reward. The bees correctly discriminated between the old-infected and novel-healthy samples, which provided confirmation of the ability of the bees to recognise specific VOCs, but not to generalize over different infected samples. In contrast to protocol 1, protocol 2 resulted in a better discrimination ability between novel infected and healthy samples at both 1 h and 24 h retention, indicating that this protocol is more efficient for training bees for SARS-CoV-2 diagnostic testing. The differences between the two protocols' outputs may be the consequence of the

bees' tendency to increase their attention during a learning task when faced with a potentially negative outcome (Avarguès-Weber et al., 2010; Chittka et al., 2003). In addition, both protocols differed in the number of conditioning rounds which could yield in a better memory. Finally, different samples were used in the conditioning phase of protocol 2, thus promoting generalization of response based on the common properties of all infected samples rather than on individual differences.

Conditioning of bees to SARS-CoV-2 derived VOCs could thus be further improved by focusing on the protocol that best worked in this study and add other elements that can make conditioning even more effective. Such an addition could be an extra training few hours after the original one or a different number of trials and alternative US. In our experiments we used appetitive–aversive conditioning due to the complexity and similarity of the odours the bees were trained for. In some cases, especially when bees are trained to fewer complex odours, the addition of a negative reinforcement can lead to lower discrimination and higher false positives (Aguiar et al., 2018)

Our results show that using single bees for diagnosis would have limited sensitivity and specificity, since the retention tests for protocol 2 showed that at 24 h post conditioning 67% of the bees correctly identified the infected sample (sensitivity) and 58% the healthy sample (specificity). A possible approach to improve diagnostic performance would be the use of multiple bees probing the same sample in parallel. In this case, diagnosis would be based on a defined number of bees (known as diagnostic threshold) reacting (expressing PER) to the sample being tested. We assessed

such an approach by performing simulations where groups of 10 bees would be used to test a sample and identified that at least six bees would have to show PER for the sample to be considered positive. By taking this approach, the potential sensitivity of the test was predicted to be around 92% and the specificity around 86%. The current standard for laboratory diagnosis of active SARS-CoV-2 infection is the detection of viral RNA from respiratory specimens by real-time, reverse transcription polymerase chain reaction (qRT-PCR). Our predicted results on accuracy are comparable to the diagnostic performance of point of care (POC) tests such as RT-LAMP tests (without RNA extraction) and rapid antigen tests. These tests showed sensitivities higher than 70% for samples with $Ct < 33$ or taken within the first week of symptom onset and specificities higher than 90% (Fowler et al., 2021; Krüttgen et al., 2021; Dinnes et al., 2021). In general, these POC tests require more than 10 min to produce a test result, whereas bees only require a few seconds to express PER (<5 s). Dogs can also be trained to detect SARS-CoV-2 and provide results very quickly. However, dogs require much more time and resources to be trained compared with bees, and their sensitivity is lower than the honeybee test (dogs' sensitivity ranges from 65-82.6%; Eskandari et al., 2021; Jendrym et al., 2020). Moreover, dogs may be infected with SARS-CoV-2 whereas bees are not sensitive to the virus. In addition, bees can be employed in remote areas where microbiological laboratory facilities are not available. As such, it can be concluded that the honeybee test is a suitable alternative, especially in situations where resources and laboratory equipment are scarce. This establishes the bee diagnostic test as an attractive monitoring method for developing countries and remote livestock communities, thanks to its low requirements and good diagnostic efficacy. The needed laboratory equipment for conditioning the bees on site will be very limited compared to other diagnostic tests for virus infections. However, reliable positive and negative control samples will be needed, and will have to be developed and tested in specialized laboratories.

Our results suggest that honeybees could be used for SARS-CoV-2 diagnosis and could potentially be applied for diagnosis of other infectious diseases. Further research is needed in order to define the duration of their memory. It is clear that their memory is weaker 24 h after the experiments compared to 1 h after the training, which might be the result of complexity and similarity of the odours. We need to identify the crucial moment in time, in which their memory retention is compromised and further assess the performance with a wider range of Ct values. Here we only tested samples with a maximum Ct of 30 and given the limited number of samples tested, we cannot assume that the performance would be similar with higher Ct values. In addition, a formal diagnostic validation study is necessary to properly validate the diagnostic approach applied under field conditions and confirm the diagnostic potential predicted in this study. The diagnostic test proposed in this study has certain weaknesses that need to be improved. The need to use multiple bees in parallel along with the laborious process of conditioning bees manually can make the preparation of the test inefficient. In addition, the bees can only be used for testing a few samples before an extension of their memory would be observed due to the absence of reward during the tests. The bees will thus have to go through a few numbers of reactivating conditioning rounds before being again operational for testing.

Conclusion

Our results indicate that the VOC profile differs between healthy and SARS-CoV2 infected minks and that honeybees can recognise these differences and discriminate between them. This performance

suggests the presence of specific biomarkers, which could be explored by performing a gas chromatography/ mass spectrometry (GC/MS) analysis. Our experiments demonstrate that bees can effectively detect the presence of an infection in samples of an extensive range of Ct values. Once improved, a diagnostic test utilizing the learning abilities of honeybees might thus provide an important addition to the current monitoring system of zoonotic diseases in remote livestock farming systems.

MATERIALS AND METHODS

Honeybee preparation

At the start of each experimental day during April and May 2021, we collected a new batch of honeybees (*Apis mellifera*) from the same beehive, located 2 km away from the Wageningen Bioveterinary Research (WBVR) laboratory in Lelystad, the Netherlands. We assumed that the bees were a mixture of different working classes. Foragers were preferred but the weather conditions did not allow for flights every day so discrimination between worker classes was not always possible. The bees were collected with a brush from inside the hive or by collecting departing bees at the hive entrance, using the same brush. For transport to the laboratory, bees were placed in transparent cylindrical plastic containers (100 ml), which carried 5-15 bees each. A total of 149 bees were used during the experiments.

The containers with honeybee workers were placed in a freezer (-20°C) for 3-5 min until the bees become inactive, which makes harnessing safer. Once out of the freezer, the bees were placed on a paper towel and inserted inside our custom-made 'bee-holders' with the help of tweezers (Fig. 3). The bee-holders are made of plastic and have the following dimensions: $20 \times 10 \times 10$ mm. They consist of two parts, a back and base, which allows the experimenter to hold it easily, and the front part, the chamber, where the bee is kept. The chamber has two openings, one in the bottom to allow for the bee to be inserted easily and a door-like structure above. The door closes once the bee is inside the chamber locking its head into position, while allowing the rest of the body to move freely. The chamber also has two openings for the bee's wings, avoiding unnecessary injuries. We harnessed the bees 30 min after collection and the experiments started 3 h after harnessing. We collected and harnessed multiple bees in parallel. Those that exhibited a PER after a brief touch of the antenna with the sugar-water solution (Fig. 3), were used for conditioning.

Sample selection

Throat swabs were taken of necropsied animals from a mink farm during the SARS-CoV2 epidemic in the period of April-November 2020, in the Netherlands. 2 ml of Dulbecco's Modified Eagle Medium (DMEM), supplemented with 10% Fetal Calf Serum (FCS) and 1% Antibiotic



Fig. 3. Picture of the conditioning procedure during protocol 1. A single honeybee harnessed inside our custom-made bee holder. The bee has just been exposed to a positive sample and been provided with a wooden stick soaked in sugar water, which has led her to express the PER.

response (PER: 0 or 1) was set as the dependent variable, while the sample (positive or negative), the conditioning round and the interaction between samples and conditioning round were fitted as fixed explanatory variables. The bees' individual identification was introduced as random intercept to account for multiple measures being made with each bee. Significance of the explanatory variables was assessed using the Wald test, with threshold for significance set to $P < 0.05$. Using these models, we were interested in assessing the improvement in the discrimination ability of bees between infected and healthy samples as a function of the number of conditioning rounds.

To assess the bee's discrimination accuracy between healthy and infected samples at 1 h and 24 h after conditioning, we fitted again a GLMM with a binomial distribution. Models were fitted for each training protocol and for each retention time (1 h or 24 h) independently. In these models, the bees' response (PER: 0 or 1) was the dependent variable, while the type of sample (new-healthy, old-healthy, new-infected, old-infected) was the fixed explanatory variable and the bees' identification was introduced as random intercept. For statistical comparison between sample types, we used the new-healthy sample as reference. This sample was used as reference because we considered that if the bees were to be used for diagnostic purpose they will be exposed to unknown (new) samples which they need to classify (discriminate between) as healthy or infected. Significance of the explanatory variables was assessed using the Wald test.

To explore the diagnostic potential of the bees and predict the diagnostic performance of the practical application of using bees for diagnosis of SARS-CoV-2 we first tested the association between the sample's Ct value (indicator of virus concentration in the sample) and detection rates after 1 h of retention. Infected samples used had Cts of 21, 27 and 30. The proportions of bees reacting to each of these samples were compared using a Chi square test. For this analysis, independence was assumed and a Bonferroni correction for multiple comparisons was applied for the interpretation of significance.

Second, we simulated a population of infected and non-infected samples which were individually tested by a group of bees. This simulation was done by random sampling with replacement groups of 10 bees, which would be part of a diagnostic group, from the retention tests done at 1 and 24 h. Sampling was done for positive samples or negative samples independently. A total of 300 groups of 10 bees exposed to positive samples and 300 groups exposed to negative samples were simulated. Sampling was done from the dataset with the retention results, which contained diagnostic results at individual bee level. From each sampled group the number of bees performing a correct discrimination of the sample (either positive or negative) was recorded. This number was then used to perform an ROC analysis to identify a potential diagnostic threshold and assess the diagnostic efficacy (sensitivity and specificity) of the system (groups of bees).

All data analyses were performed using the statistical software R version 4.0.2 (R Core Team, 2020). The GLMM were fit using the library lme4 (Bates et al., 2014). ROC analyses were done using the libraries pROC and ROCR (Robin et al., 2011; Sing et al., 2005).

Acknowledgements

We thank Bregtje Smid for outstanding laboratory technical assistance. Dr Hans Smid from the Laboratory of Entomology of Wageningen University, for his valuable insights and guidance, Saptarshi Mukhopadhyay from InsectSense B.V. and Edmund Taylor for their assistance and support throughout the project.

Competing interests

The authors declare no competing or financial interests.

Author contributions

Conceptualization: A.S., A.A., A.H., J.G., W.H.V.; Methodology: A.S., R.W.H., J.P., J.G., W.H.V.; Validation: J.G., W.H.V.; Formal analysis: J.G.; Investigation: E.K., A.S., R.W.H., J.P.; Data curation: J.G.; Writing - original draft: E.K., J.G.; Writing - review & editing: E.K., A.S., A.A., A.H., M.D., J.G., W.H.V.; Visualization: A.S., J.G.; Supervision: W.H.V.; Project administration: W.H.V.; Funding acquisition: W.H.V.

Funding

Open Access funding provided by Wageningen University. Deposited in PMC for immediate release. This study was supported by the Wageningen University and

Research KB37 research program and by funding from the European Union's Horizon 2020 Research and Innovation programme under grant agreement No 773830: One Health European Joint Programme, internal joint integrative project COVRIN.

References

- Aguilar, J., Roselino, A. C., Sazima, M. and Giurfa, M. (2018). Can honey bees discriminate between floral-fragrance isomers? *J. Exp. Biol.* **221**, jeb180844. doi:10.1242/jeb.180844
- Avarguès-Weber, A., De Brito Sanchez, M. G., Giurfa, M. and Dyer, A. G. (2010). Aversive reinforcement improves visual discrimination learning in free-flying honeybees. *PLoS ONE* **5**, e15370. doi:10.1371/journal.pone.0015370
- Bates, D., Mächler, M., Bolker, B. and Walker, S. (2014). Fitting Linear Mixed-Effects Models using lme4. Available: <https://ui.adsabs.harvard.edu/abs/2014arXiv1406.5823B> [Accessed June 01, 2014].
- Buljubasic, F. and Buchbauer, G. (2015). The scent of human diseases: a review on specific volatile organic compounds as diagnostic biomarkers. *Flavour Fragr. J.* **30**, 5-25. doi:10.1002/ffj.3219
- Chittka, L., Dyer, A. G., Bock, F. and Dornhaus, A. (2003). Psychophysics: bees trade off foraging speed for accuracy. *Nature* **424**, 388. doi:10.1038/424388a
- Control, ., E. C. F. D. P. A. (2020). *COVID-19 Testing Strategies and Objectives*. Stockholm: ECDC.
- Corman, V. M., Landt, O., Kaiser, M., Molenkamp, R., Meijer, A., Chu, D. K., Bleicker, T., Brünink, S., Schneider, J., Schmidt, M. L. et al. (2020). Detection of 2019 novel coronavirus (2019-nCoV) by real-time RT-PCR. *Euro Surveill.* **25**, 2000045.
- Dinnes, J., Deeks, J. J., Berhane, S., Taylor, M., Adriano, A., Davenport, C., Dittich, S., Emperador, D., Takwoingi, Y., Cunningham, J. et al. (2021). Rapid, point-of-care antigen and molecular-based tests for diagnosis of SARS-CoV-2 infection. *Cochrane Database Syst. Rev.* **3**, Cd013705.
- Eskandari, E., Ahmadi Marzaleh, M., Roudgari, H., Hamidi Farahani, R., Nezami-Asl, A., Laripour, R., Aliyazdi, H., Dabbagh Moghaddam, A., Zibaseresht, R., Akbarialiabad, H. et al. (2021). Sniffer dogs as a screening/diagnostic tool for COVID-19: a proof of concept study. *BMC Infect. Dis.* **21**, 243. doi:10.1186/s12879-021-05939-6
- Fitzgerald, J. E., Bui, E. T. H., Simon, N. M. and Fenniri, H. (2017). Artificial Nose Technology: Status and Prospects in Diagnostics. *Trends Biotechnol.* **35**, 33-42. doi:10.1016/j.tibtech.2016.08.005
- Fowler, V. L., Armson, B., Gonzales, J. L., Wise, E. L., Howson, E. L. A., Vincent-Mistiaen, Z., Fouch, S., Maltby, C. J., Grippon, S., Munro, S. et al. (2021). A highly effective reverse-transcription loop-mediated isothermal amplification (RT-LAMP) assay for the rapid detection of SARS-CoV-2 infection. *J. Infect.* **82**, 117-125. doi:10.1016/j.jinf.2020.10.039
- Jendryn, P., Schulz, C., Twele, F., Meller, S., Von Köckritz-Blickwede, M., Osterhaus, A., Ebberts, J., Pilchová, V., Pink, I., Welte, T. et al. (2020). Scent dog identification of samples from COVID-19 patients - a pilot study. *BMC Infect. Dis.* **20**, 536. doi:10.1186/s12879-020-05281-3
- Krüttgen, A., Cornelissen, C. G., Dreher, M., Hornef, M. W., Imöhl, M. and Kleines, M. (2021). Comparison of the SARS-CoV-2 Rapid antigen test to the real star Sars-CoV-2 RT PCR kit. *J. Virol. Methods* **288**, 114024. doi:10.1016/j.jviromet.2020.114024
- Matsumoto, Y., Menzel, R., Sandoz, J. C. and Giurfa, M. (2012). Revisiting olfactory classical conditioning of the proboscis extension response in honey bees: a step toward standardized procedures. *J. Neurosci. Methods* **211**, 159-167. doi:10.1016/j.jneumeth.2012.08.018
- Olsson, M. J., Lundström, J. N., Kimball, B. A., Gordon, A. R., Karshikoff, B., Hosseini, N., Sorjonen, K., Olgart Höglund, C., Solares, C., Soop, A. et al. (2014). The scent of disease: human body odor contains an early chemosensory cue of sickness. *Psychol. Sci.* **25**, 817-823. doi:10.1177/0956797613515681
- Oreshkova, N., Molenaar, R. J., Vreman, S., Harders, F., Oude Munnink, B. B., Hakke-Van Der Honing, R. W., Gerhards, N., Tolsma, P., Bouwstra, R., Sikkema, R. S. et al. (2020). SARS-CoV-2 infection in farmed minks, the Netherlands, April and May 2020. *Euro Surveill.* **25**, 2001005. doi:10.2807/1560-7917.ES.2020.25.23.2001005
- Pavlov, P. I. (1927). Conditioned reflexes: an investigation of the physiological activity of the cerebral cortex. *Ann Neurosci* **17**, 136-141.
- Probert, C. S., Ahmed, I., Khalid, T., Johnson, E., Smith, S. and Ratcliffe, N. (2009). Volatile organic compounds as diagnostic biomarkers in gastrointestinal and liver diseases. *J. Gastrointest. Liver Dis.* **18**, 337-343.
- R Core Team (2020). *A Language and Environment for Statistical Computing, Version 4.0.2*. ed. Vienna, Austria: R Foundation for Statistical Computing.
- Robin, X., Turck, N., Hainard, A., Tiberti, N., Lisacek, F., Sanchez, J. C. and Müller, M. (2011). pROC: an open-source package for R and S+ to analyze and compare ROC curves. *BMC Bioinformatics* **12**, 77. doi:10.1186/1471-2105-12-77
- Sethi, S., Nanda, R. and Chakraborty, T. (2013). Clinical application of volatile organic compound analysis for detecting infectious diseases. *Clin. Microbiol. Rev.* **26**, 462-475. doi:10.1128/CMR.00020-13

- Shirasu, M. and Touhara, K.** (2011). The scent of disease: volatile organic compounds of the human body related to disease and disorder. *J. Biochem.* **150**, 257-266. doi:10.1093/jb/mvr090
- Sing, T., Sander, O., Beerenwinkel, N. and Lengauer, T.** (2005). ROCR: visualizing classifier performance in R. *Bioinformatics* **21**, 3940-3941. doi:10.1093/bioinformatics/bti623
- Strauch, M., Lüdke, A., Münch, D., Laudes, T., Galizia, C. G., Martinelli, E., Lavra, L., Paolesse, R., Ulivieri, A., Catini, A. et al.** (2014). More than apples and oranges—detecting cancer with a fruit fly's antenna. *Sci. Rep.* **4**, 3576. doi:10.1038/srep03576
- Suckling, D. M. and Sagar, R. L.** (2011). Honeybees *Apis mellifera* can detect the scent of *Mycobacterium tuberculosis*. *Tuberculosis (Edinb.)* **91**, 327-328. doi:10.1016/j.tube.2011.04.008
- Sutherland, A. M., Gubler, W. D., Wingo, R. M. and McCabe, K. J.** (2010). Classical conditioning of domestic honeybees to olfactory stimuli associated with grapevine powdery mildew infections. Proceedings of the 6th International Workshop on Grapevine Downy and Powdery Mildew. Bordeaux, France.
- Takeda, K.** (1961). Classical conditioned response in the honey bee. *J. Insect Physiol.* **6**, 168-179. doi:10.1016/0022-1910(61)90060-9
- Trabue, S., Scoggin, K., Li, H., Burns, R., Xin, H. and Hatfield, J.** (2010). Speciation of volatile organic compounds from poultry production. *Atmos. Environ.* **44**, 3538-3546. doi:10.1016/j.atmosenv.2010.06.009
- Wells, C. R., Townsend, J. P., Pandey, A., Moghadas, S. M., Krieger, G., Singer, B., McDonald, R. H., Fitzpatrick, M. C. and Galvani, A. P.** (2021). Optimal COVID-19 quarantine and testing strategies. *Nat. Commun.* **12**, 356. doi:10.1038/s41467-020-20742-8
- Wilson, A. D.** (2018). Applications of electronic-nose technologies for noninvasive early detection of plant, animal and human diseases. *Chemosensors* **6**, 45. doi:10.3390/chemosensors6040045
- Wright, G. A., Mustard, J. A., Simcock, N. K., Ross-Taylor, A. A., McNicholas, L. D., Popescu, A. and Marion-Poll, F.** (2010). Parallel reinforcement pathways for conditioned food aversions in the honeybee. *Curr. Biol.* **20**, 2234-2240. doi:10.1016/j.cub.2010.11.040