

https://dspace.org/

New volatile organic compounds from the exhaled breath of active tuberculosis patients

Atlang Gild. Mpolokang; Tlotlo Cassandra. Setlhare.; Somnath Bhattacharyya; George Chimowa

https://hdl.handle.net/10539/46156
WOS

Downloaded from a DSpace repository. https://dspace.org/

scientific reports

Check for updates

OPEN

New volatile organic compounds from the exhaled breath of active tuberculosis patients

Atlang Gild. Mpolokang[1], Tlotlo Cassandra. Setlhare[1], Somnath Bhattacharyya[2] & George Chimowa[1]✉

Tuberculosis (TB) remains a leading global cause of mortality, necessitating the development of innovative diagnostic approaches for early and accurate detection. Existing diagnostic methods are invasive, time-intensive, and contribute to delays in diagnosis, thereby exacerbating disease progression in patients and facilitating community transmission. To address these limitations, this study investigated human breath samples to identify volatile organic compounds (VOCs) associated with active TB as potential non-invasive biomarkers. VOCs were analysed using gas chromatography-mass spectrometry (GC-MS), supported by AMDIS and OpenChrom software for compound identification. The findings revealed the presence of previously unreported VOCs associated to breath samples from patients with active TB and multidrug-resistant TB (MDR-TB), which were absent in control participants without TB symptoms. Furthermore, the results suggest the feasibility of differentiating MDR-TB from active TB based on breath VOC profiles, marking a novel observation. These findings underpin the development of a non-invasive breathaliser diagnostic device for TB, with the potential to enhance disease management and reduce transmission.

Keywords Tuberculosis, Human breath, Volatile organic compounds, VOCs profiling

Tuberculosis (TB) remains a critical global health challenge, with approximately 10 million new cases and 1.4 million deaths reported annually. These figures are particularly alarming in southern Africa, where this study was conducted, and in other low-income regions, as highlighted by the World Health Organization in 2023[1]. TB is a pulmonary disease that ranks among the most lethal infectious diseases worldwide, surpassing the human immunodeficiency virus (HIV) and, in some regions, Coronavirus Disease 2019 (COVID-19)[1,2]. A more severe manifestation of TB, multidrug-resistant tuberculosis (MDR-TB), presents heightened challenges due to limited treatment options and elevated mortality rates compared to drug-susceptible TB[3]. MDR-TB arises when Mycobacterium tuberculosis develops resistance to rifampicin and isoniazid, the two most potent anti-TB drugs[4]. This issue is exacerbated in southern Africa by the high prevalence of HIV and AIDS, as people with compromised immune systems are more susceptible to primary transmitted resistance than to acquired resistance[5]. This interplay complicates the management and treatment of MDR-TB, underscoring the urgent need for innovative, low-cost, and rapid diagnostic tools. Early detection of active TB is essential for effective disease management, particularly in densely populated urban areas where informal settlements are expanding due to rural-to-urban migration. Performing a timely diagnosis can significantly reduce transmission rates within such communities.

A promising approach to TB diagnostics involves analysing volatile organic compounds (VOCs) present in exhaled human breath and it may also be used for treatment monitoring[6,7]. Current diagnostic methods, such as culture and smear microscopy, often rely on sputum samples, which some patients-particularly children and those critically ill-may be unable to produce. While the introduction of Xpert MTB/RIF has addressed some limitations of smear microscopy, challenges persist, including poor adherence to diagnostic algorithms, high costs, and supply chain disruptions for cartridges. Furthermore, existing diagnostic methods can take several weeks to yield results, delaying timely intervention and allowing disease progression and transmission[8,9]. These challenges explain why significant research efforts have been conducted to explore the potential of volatile organic compounds (VOCs) present in human breath as possible biomarkers for active TB detection[10–17]. The metabolic and oxidative stress responses triggered by active *M. tuberculosis* infection result in the production of VOCs, some of which may serve as distinctive biomarkers for TB. The ability to quantitatively and qualitatively

[1]Department of Physics and Astronomy, Botswana International University of Science and Technology, Private Bag 16, Palapye, Botswana. [2]Nanoscale Transport Physics Laboratory, School of Physics, University of the Witwatersrand, Wits, P. Box 2, Johannesburg 2050, South Africa. ✉email: chimowag@biust.ac.bw

measure these VOCs offers a non-invasive, rapid and potentially cost-effective approach to TB diagnosis. By identifying specific VOCs associated with active TB and MDR-TB, novel diagnostic devices can be developed[18], paving the way for early detection and disease monitoring. Previous studies employing gas chromatography-mass spectrometry (GC-MS) and various software tools, such as ChromaTOF[15], Xcalibur Qual Browser[16], Python[19] and R packages[20], have identified potential biomarkers associated with active tuberculosis. Machine learning algorithms, including support vector machines, principal component analysis, and random forests, just to name a few, have been utilised to classify TB status based on VOC profiles. However, these approaches remain constrained by the reliance on laboratory-based GC-MS data, limiting their practicality for rapid diagnosis. A list of compounds identified from human breath and TB cultures reported as distinctive VOCs of TB from different studies are shown in table 1.

While significant progress has been made, there is still no standardised technique for diagnosing active TB via breath analysis. One potential limitation is the focus on identifying individual VOCs specific to TB. Given that metabolic responses to infection may produce similar VOCs across different diseases, this approach may not be universally effective. Instead, analysing the complete VOC profile and relative concentrations associated with TB could enable the differentiation of diseases with overlapping VOCs.

This study focuses on profiling VOCs associated with active TB and MDR-TB using a single quadrupole GC-MS[22] coupled with OpenChrom[23] and Automated Mass Spectral Deconvolution and Identification System (AMDIS) software[24]. The results report, for the first time, novel VOCs linked to these conditions, with the long-term aim of developing sensitive and selective gas-sensing materials for diagnostic electronic nose (E-nose) devices. Such devices could revolutionise TB diagnostics by providing rapid, non-invasive, and cost-effective point-of-care solutions. While this work primarily addresses the profiling phase, it sets the foundation for subsequent development of these innovative diagnostic tools.

Breath analysis as a diagnostic modality has long been recognised but remains largely experimental due to technological limitations in detecting trace-level VOCs. The advent of nanotechnology and AI-based technologies offers renewed optimism in this field, enabling the precise identification of VOCs at concentrations relevant for disease diagnostics. The advantages of breath analysis, non-invasiveness, rapidity, and affordability, position it as a transformative tool for TB diagnosis and monitoring.

Results

Human breath samples collected from patients with active tuberculosis (TB) and control participants (individuals without TB symptoms) were analysed using gas chromatography-mass spectrometry (GC-MS). The samples were injected into the GC-MS system, where helium gas served as the carrier, transporting volatile organic compounds (VOCs) through a capillary column. Interaction between the VOCs and the stationary phase of the column resulted in separation based on the volatility of the compounds. Highly volatile compounds exhibited lower retention times due to faster interaction with the stationary phase, whereas less volatile compounds demonstrated higher retention times.

Representative chromatograms of breath samples from a TB patient and a control participant are shown in Figure 1. Distinct peaks, observed in the chromatogram of the TB sample (red trace), are absent in the control sample (green trace), indicating notable differences in the VOC composition between the two groups. These findings suggest the presence of VOCs that can be associated to active TB, underscoring their potential as possible biomarkers for disease detection.

Analysis of the breath samples using OpenChrom software identified an average of 825 compounds per sample. Following the exclusion of contaminants observed in blank runs and compounds associated with Tedlar bags, such as N,N-dimethylacetamide (DMAC), phenol, and acetic acid[25], a total of 1,397 features were extracted.

Sample type	Compounds	Analysis Technique	Ref
Breath	Naphthalene 2,3,6-trimethyl-, Heptane 4-ethyl-2,2,6,6-tetramethyl-, Indane, 2,2,3-trimethylhexane, 1-Hexene-6-phenyl-4-(1-phenylethoxy)-, Butyl acetate, 1,2,3-trimethylbenzene, 2,2,4,6,6-pentamethylheptane, 4-(2-methylbutan-2-yl) phenol, Ethyl butyrate, 4-methyl-1-decene	GC + GC-TOFMS	16
Breath	Decane, 4-methyloctane	GC + GC-TOFMS	20
Breath	1,4-Dichlorobenzene, Trans-1,3-Dimethylcyclohexane, 1,4-Dimethylcyclohexane, 1-Octanol 2-butyl, 1-Methylnaphthalene, Camphene, 4-Methyldecane, 3-Ethyl-2-methylheptane, 2,6-Dimethyloctane, 1,2,3,4-Tetramethylbenzene,Bicyclo 3.1.1 hept-2-ene, 3,6,6-trimethyl-, trans-1-Ethyl-4-Methlcylohexane.	ATD-GC-MS,	10
Breath	Oxetane 3-(1-methylethyl)-, 4-Methyldodecane, Hexylcyclohexane, Bis-(3,5,5-trimethylhexyl) phthalate, 1,3,5-trimethylbenzene, 3,7-Dimethyldecane, Tridecane, 4,6,8-Trimethyl-1-nonene, 5-Ethyl-2-methylheptane, 4-Methyl-1-hexene	ATD-GC-MS	11
Culture	1-Methylnaphthalene, 3-heptanone, 2,2,4,6,6-Pentamethylheptane, Benzene 1-methyl-4-(1-methylethyl)-, 1,4-Dimethylcyclohexane, 3,5-dimethylamphetamine, Methylcyclododecane, 3-Methylbutanal, 2-hexene, Trans-anti-1-methyldecahydronaphthalene	ATP-GC-MS	10
Culture	Hexadecanoic acid, Decane, Dodecane, Tridecane, 2-Methyltridecane, tridecanol, Octadecanoic acid, Isotridecanol, Eicosonoic acid, Nonadecanoic acid	GC + GC-TOFMS	21
Culture	Cyclohexanone 2-(dimethylamino)-, 2-propanamine, isobutyronitrile, ethane, isocyanato-, dodecane, cyclohexanone, 3-pentene-2-one, Methanethioanide N N-dimethyl-, Ethanamine N-methyl	GC-TOF-MS	13
Culture	4-ethyl-2,2,6,6-tetramethlheptane, 2-ethylhexyl isobutyl sulfite, 3-hydroxy-3-methylbutanoic acid, 2-Pinene, Azulene, Heptanal, 1-propynylbenzene, 2,3,6-trimethylheptane, 1,2-xylene, 4-ethyl decane, 2,5-dimethyldecane, 1-nitroadamantane, 2,6-dimethylheptadecane, beta-phellandrene, 4-tert-Amyl phenol	GC + GC-TOFMS	16

Table 1. A list of reported possible biomarkers of TB identified from human breath and mycobacterium culture across different studies.

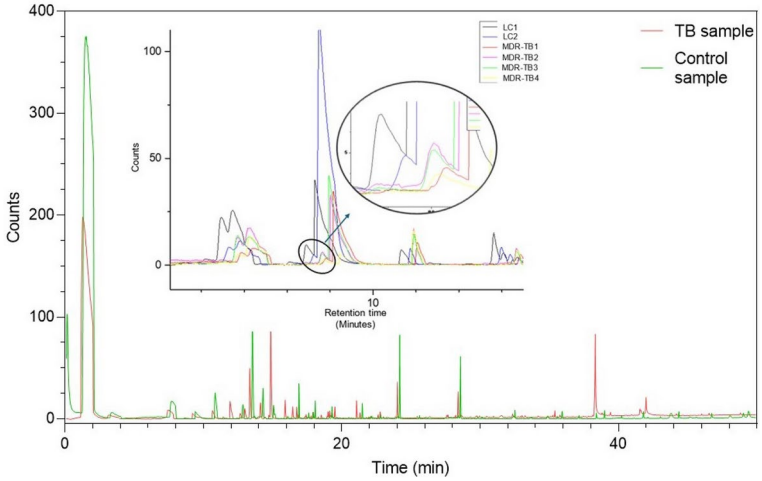

Figure 1. A typical chromatogram of TB and control breath samples shows multiple peaks at different retention times, with some peaks only in the TB patient and not in the control Sample. Inset: Graph showing peaks of 4-methyloctane in lung cancer (LC) patients and MDR-TB subjects. LC patients have higher concentrations of the VOC, when compared to the TB patients.

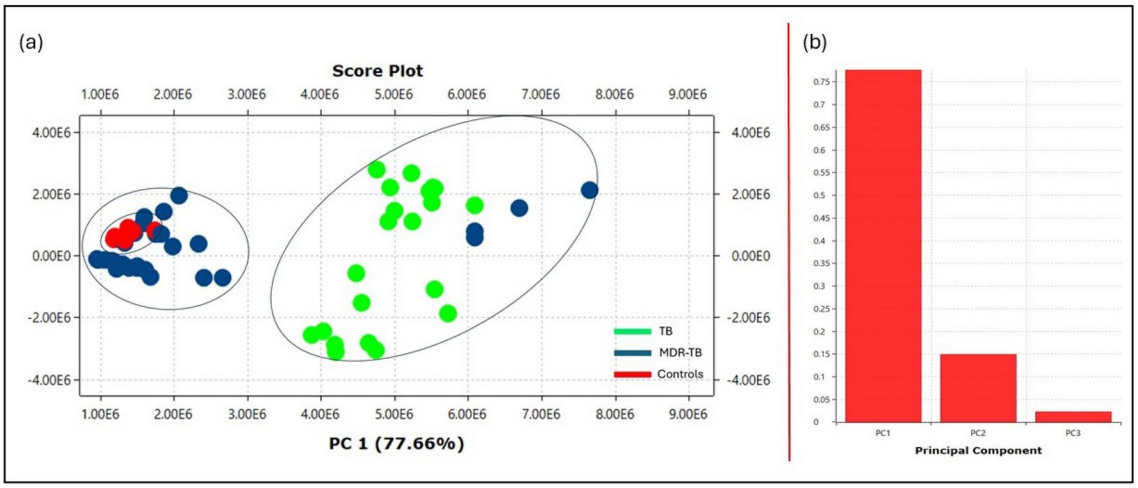

Figure 2. (**a**) Principal Component Analysis score plot. Red represents controls, blue are MDR-TB subjects, and green are TB subjects. (**b**) Scree bar plot showing the variance of individual principal components, from this bar graph PC1 and PC2 are taken as the cut-off.

Principal component analysis (PCA), a statistical method for dimensionality reduction and data visualisation, was applied to the extracted features to explore the segregation between participants with active TB, multidrug-resistant TB (MDR-TB), and controls.

PCA enables the segregation of data based on the characteristic peaks (as indicated by retention times) and concentration levels of the identified VOCs. The principal component score plot, Figure 2(a), reveals distinct clustering: Active TB subjects form a separate cluster, clearly segregated from both controls and MDR-TB subjects, while MDR-TB participants are clustered together but separate from active TB patients. This indicates that active TB, MDR-TB, and control groups exhibit distinct VOC profiles. However, MDR-TB participants show some overlap with controls, suggesting limited variation between these groups.

This observation aligns with findings by Chen et al.[19], who applied PCA to features extracted from high-resolution mass spectrometry data and observed a similar trend, with non-TB subjects forming a tightly clustered group and a discernible separation between TB and non-TB participants. Additionally, a small subset of MDR-TB participants overlaps with the TB group. This overlap is attributable to shared compounds such as tridecane, decane, and o-cymene, though MDR-TB samples differ in having higher concentrations of tridecane compared to TB samples.

The scree bar plot Figure 2(b) illustrates the variance explained by individual principal components. Principal components 1 (PC1) and 2 (PC2) account for 90% of the data variance and were therefore selected as the cut-off

for analysis. This variance is primarily determined by the concentration of VOCs and their retention times, with PC1 representing the primary source of variation in the dataset.

A total of eight volatile organic compounds (VOCs) were identified as contributors to the variation observed between active TB, MDR-TB, and control groups (see, Table 2). Among these, seven VOCs-tridecane, decane, o-cymene, 4-methyloctane, L-β-pinene, 1-octanol-2-butyl, and dodecane-4-methyl-were associated with active TB, while benzene-1-methyl-4-(1-propynyl) emerged as a potential indicator of MDR-TB.

Tridecane, decane, 4-methyloctane, L-β-pinene, 1-octanol-2-butyl, and dodecane-4-methyl have been previously reported as candidate biomarkers for TB in Table 1. However, to the best of our knowledge, this study is the first to identify o-cymene as a novel VOC associated with active TB, detected in 25% of the normal TB patient cohort and in 27% of the MDR-TB cohort, giving a 52% prevalance in active TB patients. Additionally, benzene-1-methyl-4-(1-propynyl) was identified in 45% of MDR-TB patients, making it a potential biomarker for MDR-TB, as shown in Table 2. Importantly, none of these VOCs were detected in control participants, supporting their potential as disease-specific biomarkers when used in combination with other VOCs.

We propose that accurate diagnosis of active TB or MDR-TB using human breath should not rely on the identification of a single VOC. Instead, a combination of specific VOCs must be considered, reflecting the complexity of metabolic processes and ensuring greater diagnostic accuracy, as previously discussed.

Discussions

The chromatogram peaks and mass spectra for the newly identified VOCs, o-cymene and benzene-1-methyl-4-(1-propynyl), detected in active TB and MDR-TB breath samples respectively, are presented in Figure 3(a -d). The mass spectrum of o-cymene demonstrated a match factor of 89.0/100 and a reverse match factor of 89.4/100 against the NIST library, while benzene-1-methyl-4-(1-propynyl) exhibited a match factor of 83.2/100 and a reverse match factor of 84.5/100. Both values align with the NIST guidelines[26] for good and acceptable matches.

O-cymene, also referred to as benzene-1-methyl-2-(1-methylethyl), shares structural similarity with m-cymene (benzene-1-methyl-4-(1-methylethyl)), a compound previously identified from Mycobacterium tuberculosis cultures in a study by Phillips et al.[10]. This similarity is attributed to the positional difference of the 1-methylethyl group, suggesting a common source for these VOCs.

Notably, both o-cymene and benzene-1-methyl-4-(1-propynyl) are benzene derivatives, indicating their potential origin from metabolic products of *M. tuberculosis* or from complex amino acids metabolised by the host[27]. These findings contribute to the growing evidence of the utility of VOC profiling in distinguishing between active TB and MDR-TB, offering new insights into potential disease-specific biomarkers.

Furthermore, benzene-1-methyl-4-(1-propynyl), uniquely detected in MDR-TB patients, shares structural similarities with o-cymene, which was identified in active TB patients. This suggests that o-cymene undergoes modification to form benzene-1-methyl-4-(1-propynyl) in MDR-TB cases, as illustrated in Figure 3(e). A plausible explanation for this transformation is that, as Mycobacterium tuberculosis develops drug resistance[28], it mutates, resulting in altered metabolic pathways. These mutations, which enable bacterial survival in the

Name of VOC	Retention Time	Mass/Charge	CAS No	No of subjects		
				TB%	MDR-TB%	Controls %
Tridecane	22.80 - 23.40	184.0	629-50-5	91.7	45.5	16.7
Decane	13.85 -14.00	142.0	124-18-5	50.0	27.3	0.0
Dodecane 4-methyl	22.17 - 22.34	184.0	6117-97-1	8.3	0.0	0.0
1-Octanol 2-butyl	23.50 -24.39	186.0	3913-02-8	8.3	0.0	0.0
O-cymene	14.65 - 14.80	134.0	527-84-4	25.0	27.3	0.0
4-Methyloctane	8.89 -9.18	128.0	2216-34-4	8.3	36.4	0.0
Benzene 1-methyl 4-(1-propynyl)	17.80-18.22	130.0	2749-93-1	0.0	45.5	0.0
L-β-pinene	12.94 - 13.14	136.0	18172-67-3	16.7	0.0	0.0

Table 2. A list of compounds identified as possible biomarkers of TB and MDR-TB. The new Biomarkers reported for the first time are highlighted in red.

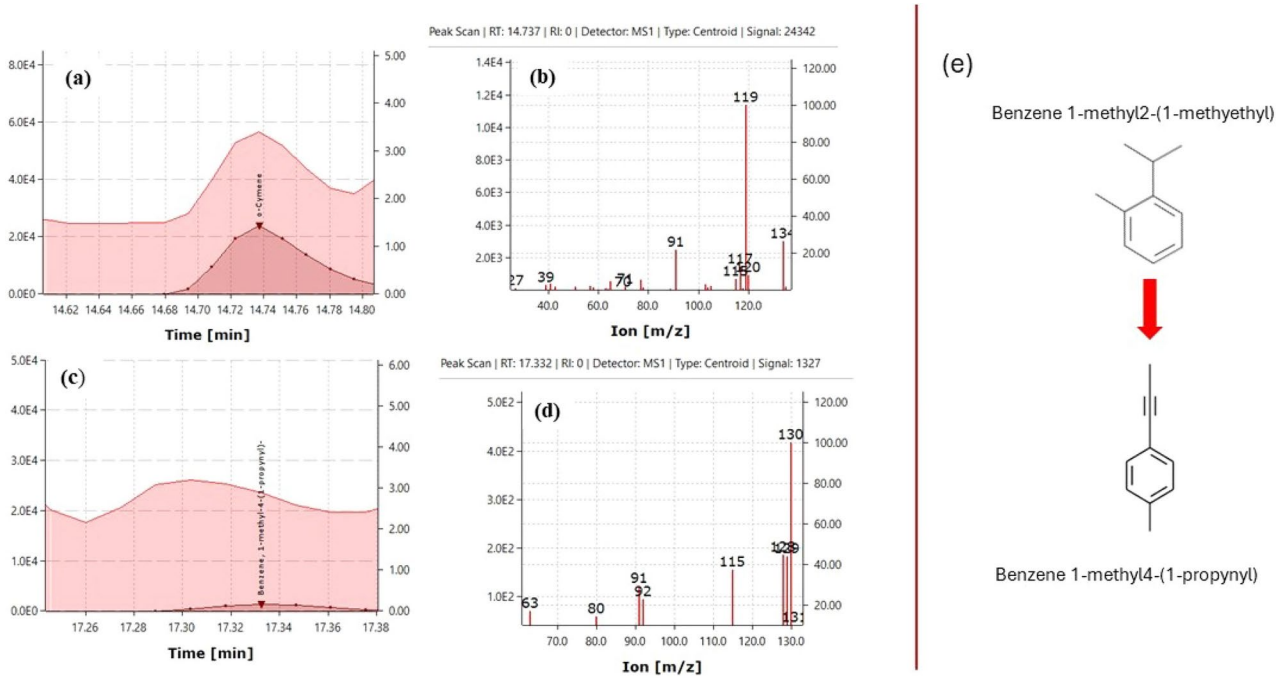

Figure 3. Chromatogram peaks of (**a**) O-cymene and (**c**) Benzene 1-methyl 4-(1-propynyl) detected from TB breath sample and MDR-TB sample and their respective mass to charge spectra (**b**) and (**d**). (**e**) Structure of Benzene 1-methyl 2-(1-methyethyl) in normal TB patients modified to Benzene 1-methyl 2-(1-propynyl) in MDR-TB patients.

presence of anti-TB drugs, may produce benzene-1-methyl-4-(1-propynyl) as a by-product of the modified metabolic processes.

This hypothesis aligns with previous studies that have linked benzene derivatives to the metabolic processes of *M. tuberculosis*[27], though further investigations are required to confirm these findings. Importantly, the observed modification of o-cymene in TB patients to benzene-1-methyl-4-(1-propynyl) in MDR-TB patients presents a promising avenue for differentiating between active TB and MDR-TB through breath analysis. This distinction represents a significant advancement in the application of breath analysis technologies for disease diagnosis.

The ability to distinguish between different strains of the same disease using non-invasive breath analysis has the potential to revolutionise diagnostic approaches for this highly infectious disease. Furthermore, we note that 4-methyloctane, previously reported by Bobak et al.[20] in a study of children with a mean age of 6 years, was also detected in adults with active TB and MDR-TB patients in the present study. Interestingly, 4-methyloctane has also been reported in adults with chronic obstructive pulmonary disease (COPD), asthma, and lung cancer (in our another on-going research work), suggesting its broader association with respiratory conditions and may not be specific to active TB. The inset of Figure 1 highlights peaks of 4-methyloctane observed in lung cancer subjects (LC1 and LC2) alongside MDR-TB subjects. Notably, lung cancer patients exhibited a higher average concentration of 4-methyloctane (3599880.5 min) compared to TB patients (1624733.7 min), with concentrations determined by the area under the peak. The presence of similar VOCs across different diseases underscores the importance of analysing a comprehensive profile of VOCs, rather than relying on individual compounds, to differentiate between conditions. This approach is particularly crucial for diseases affecting the same organ, such as the lungs, where overlapping metabolic processes may produce comparable VOCs.

As shown in Table 2, tridecane was detected across all study groups; however, it was identified in only a small subset of the control participants. Upon further analysis, the concentration of tridecane was observed to be highest in MDR-TB patients, followed by TB patients, and lowest in the control group. This observation aligns with findings by Loots et al.[21], who reported elevated tridecane concentrations in drug-resistant TB strains. These results suggest that monitoring tridecane concentration may serve as a potential indicator of TB disease progression. However, the presence of tridecane in some control subjects underscores the necessity of using a broader VOC profile for definitive diagnosis.

To validate these findings, we analyzed a standard tridecane compound using GC-MS. The chromatographic peaks of tridecane detected in TB patient breath samples matched those from the headspace of the tridecane standard, as shown in Figure 4. Both exhibited similar retention times, with match factors of 92.6/100 and 95.7/100, confirming the reliability of the analytical method employed in this study.

This investigation demonstrates clear distinctions in the breath profiles of healthy participants, active TB patients, and MDR-TB patients. Importantly, it establishes, for the first time to our knowledge, that human breath analysis can differentiate between normal active TB and MDR-TB-a significant advancement in non-invasive diagnostic approaches.

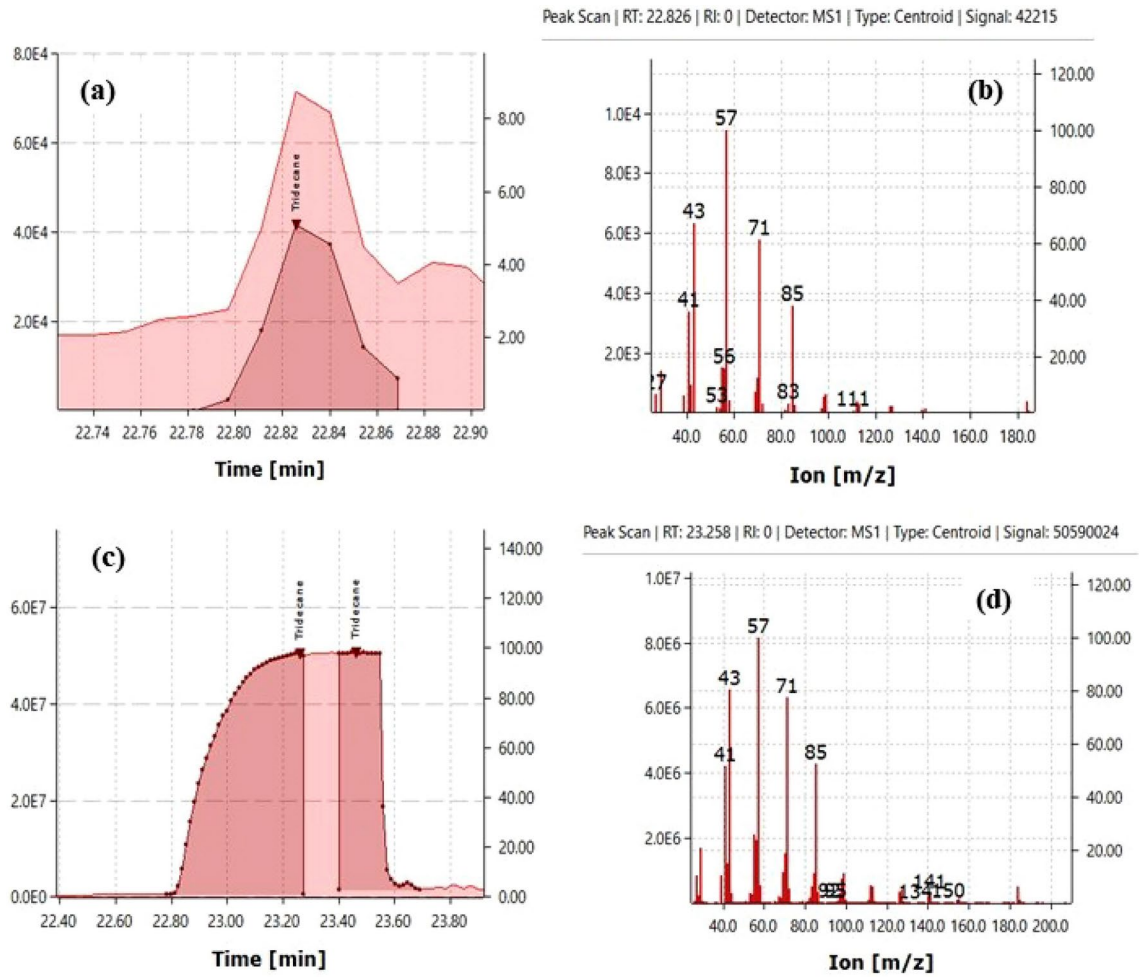

Figure 4. (**a**) Chromatogram peak of tridecane compound detected from TB breath sample and (**b**) a m/z spectrum of extracted tridecane peak showing tridecane match with spectrum from NIST MS library. (**c**) Chromatogram peak of tridecane compound detected from headspace of tridecane standard and (**d**) a m/z spectrum of extracted tridecane peak showing tridecane match with spectrum from NIST MS library.

Conclusions

In this study, we identified eight volatile organic compounds (VOCs) that distinguish active TB from multidrug-resistant TB (MDR-TB) and non-TB subjects through analysis of exhaled human breath using gas chromatography-mass spectrometry (GC-MS). Among these, we report o-cymene as a novel biomarker for TB, alongside six VOCs previously documented in the literature. Additionally, Benzene 1-methyl 4-(1-propynyl) is presented as a distinctive VOC capable of differentiating TB from MDR-TB.

Our findings further reveal that the concentration of 4-methyloctane varies significantly across pulmonary diseases, with notably higher levels in lung cancer patients compared to active TB patients, suggesting its potential utility in differentiating these conditions. This study establishes a robust foundation for the development of sensitive and selective gas sensors targeting VOCs specific to active TB. Such advancements pave the way for an electronic nose (e-nose) device, offering rapid, non-invasive, and point-of-care diagnostic capabilities.

By enabling early detection and improved treatment strategies, the proposed e-nose technology could revolutionise active TB diagnosis and management. This work represents a significant step towards innovative diagnostic solutions, contributing to the global fight against tuberculosis and its multidrug-resistant strains.

Methods

Study population and clinical characteristics

A total of 41 individuals participated in this study after providing informed consent, including, where applicable, consent from legal guardians. The cohort consisted of 12 individuals diagnosed with active pulmonary tuberculosis (TB), 11 individuals with multidrug-resistant tuberculosis (MDR-TB), and 18 control subjects. Over the course of three consecutive days, 80 breath samples were collected and analysed. The gender distribution among MDR-TB patients was balanced, with approximately 50% male and 50% female participants, see Figure 5(a). The age range across all participants spanned from 17 to 79 years, with a mean age of 34 years for active TB patients and 32 years for MDR-TB patients. TB cases were confirmed via positive sputum culture tests and/

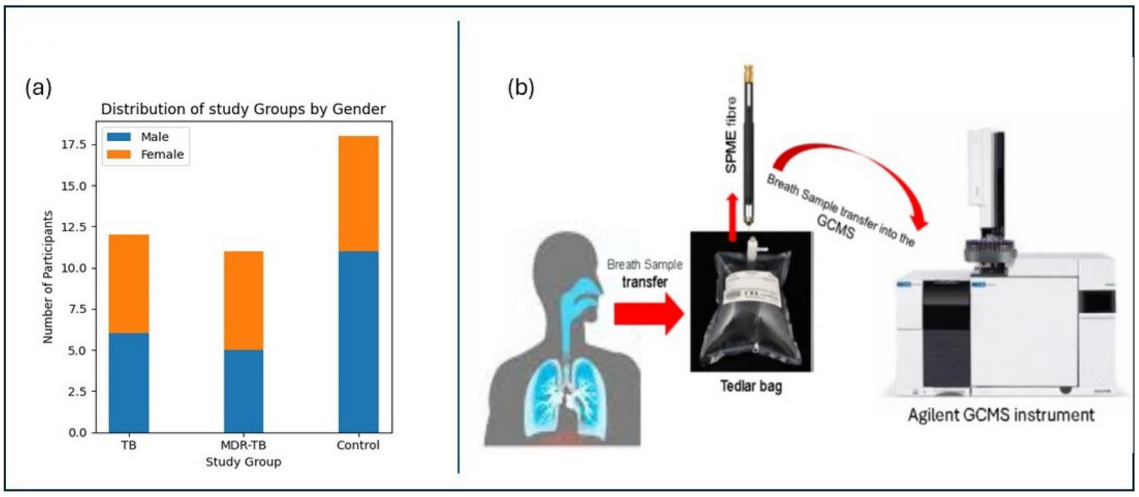

Figure 5. (**a**) Bar chart showing distribution of study groups by gender and role in the study. (**b**) Schematic diagram illustrating the process of sample collection and analysis.

or the GeneXpert MTB/RIF assay, while MDR-TB cases were diagnosed using GeneXpert MTB/RIF testing. Active TB patients were outpatients attending local clinics in Palapye and Serowe, ensuring a diverse sample population. To minimise potential confounding factors such as dietary or medicinal influences, breath samples were collected early in the morning before participants consumed food or medication. Conversely, MDR-TB samples were obtained from hospitalised patients in isolation wards in Gaborone and Francistown, Botswana. These patients had controlled diets and engaged in minimal physical activity during their hospital stay. Participation in the study was entirely voluntary. Control subjects, aged between 17 and 65 years, were included to provide a baseline for comparison. These individuals were not tested for active TB and displayed no symptoms of active TB, and thus were presumed to be free from the disease. The potential presence of latent tuberculosis (TB) in the control subjects presents a complex issue, warranting further investigation. However, the current study specifically addresses active TB, with the assumption that latent TB bacteria do not substantially influence metabolic processes or oxidative stress in affected individuals. Consequently, latent TB is unlikely to have a significant impact on the volatile organic compound (VOC) profile. This may explain why individuals with latent TB do not exhibit clinical symptoms. Based on this premise, it can be inferred that the VOC profile of latent TB patients may closely resemble that of non-TB individuals. ***Declaration***: All sample collection, handling, and data analysis procedures adhered to guidelines approved by the Ministry of Health and Wellness in Botswana and the ethics committees of the following institutions in Botswana, Human Resources development Council (HRDC), Princess Marina Hospital, Nyangabgwe Refeerral Hopsital, Palapye Hospital and Clinics. The experimental protocols followed international standards for research involving human participants namely the Declaration of Helsinki.

Sample collection and analytical procedure

Breath samples were collected using 0.5-litre Tedlar bags (Keika Ventures, USA) equipped with a mouthpiece. Participants were instructed to exhale into the bag through the mouthpiece early in the morning prior to consuming food or medication. To prevent re-inhalation of exhaled air, participants lightly pinched their noses while filling the bag. Following collection, the samples were stored in a cold environment to minimise compound decomposition during transportation from the collection sites to the laboratory, where they were analysed on the same day.

Volatile organic compounds (VOCs) were analysed using a single quadrupole GC/MS system (Agilent 7890B GC coupled with 5977A MSD). Extraction of VOCs from the Tedlar bags was performed using a solid-phase micro-extraction (SPME) fibre (2 cm, 50/30 μm Divinylbenzene/Carboxen/Polydimethylsiloxane, DVB/CAR/PDMS). The SPME fibre was inserted into the bag to pre-concentrate the compounds for several minutes, after which it was immediately transferred to the GC/MS system for analysis, see the schematic diagram, Figure 5(b).

Separation of compounds was achieved using an Agilent J and W DB-5 capillary column (30 m $\times$ 0.25 mm i.d., 0.25 μm film thickness). The oven temperature programme began at 35 °C, held for 5 minutes, followed by an increase at a rate of 5 °C/min to 250 °C, which was maintained for 2 minutes, resulting in a total run time of 50 minutes. Helium was used as the carrier gas at a flow rate of 1.2 mL/min in splitless mode. Ionisation was performed via electron impact at 70 eV, with a mass spectrometer (MS) source temperature of 230 °C. Mass spectra were acquired in the range of 25-350 m/z at a scan rate of 1.2 scans/sec with a signal threshold of 100.

Chromatographic data acquisition, peak detection, and deconvolution were conducted using OpenChrom Lablicate Edition 1.5.0 (https://www.openchrom.net) and Automated Mass Spectral Deconvolution and Identification System (AMDIS) software (https://chemdata.nist.gov/dokuwiki/doku.php?id=chemdata:amdis). Compound identification was facilitated using the NIST 11 MS library (v2.3) which came with the Instrument, applying minimum match and reverse match factors of 80/100. When necessary, analytical standards (e.g.,

tridecane) were employed for compound verification. Data classification and segregation were performed using Principal Component Analysis (PCA), ensuring robust statistical analysis of the detected VOC profiles.

Data availability

The datasets generated and/or analysed during the current study are not publicly available due to the need to protect the privacy of participants and is still also being used for future work which has not yet been published but are available from the corresponding author on reasonable request.

Received: 4 October 2024; Accepted: 3 February 2025
Published online: 12 February 2025

References

1. Chakaya, J. et al. The who global tuberculosis 2021 report-not so good news and turning the tide back to end tb. *Int. J. Infect. Dis.* **124**, S26–S29 (2022).
2. Organization, W. H. Global tuberculosis report 2018 (World Health Organization, 2018).
3. Orgnization W. H. Who consolidated guidelines on tuberculosis: Module 4: treatment - drug-resistant tuberculosis treatment, 2022 update. Internet, https://www.ncbi.nlm.nih.gov/books/NBK588564/ (2022).
4. Mosquera-Restrepo, S. F. et al. A mycobacterium tuberculosis fingerprint in human breath allows tuberculosis detection. *Nat. Commun.* **13**, 7751 (2022).
5. Shenoi, S., Heysell, S., Moll, A. & Friedland, G. Multidrug-resistant and extensively drug-resistant tuberculosis: Consequences for the global HIV community. *Curr. Opin. Infect. Dis.* **22**, 11–17 (2009).
6. Saktiawati, A. M., Putera, D. D., Setyawan, A., Mahendradhata, Y. & van der Werf, T. S. Diagnosis of tuberculosis through breath test: A systematic review. *EBioMedicine* **46**, 202–214 (2019).
7. Steingart, K. R. et al. Xpert® mtb/rif assay for pulmonary tuberculosis and rifampicin resistance in adults. Cochrane database of systematic reviews (2014).
8. Fogel, N. Tuberculosis: A disease without boundaries. *Tuberculosis* **95**, 527–531 (2015).
9. Ryu, Y. J. Diagnosis of pulmonary tuberculosis: Recent advances and diagnostic algorithms. *Tuberc. Respir. Dis.* **78**, 64 (2015).
10. Phillips, M. et al. Volatile biomarkers of pulmonary tuberculosis in the breath. *Tuberculosis* **87**, 44–52 (2007).
11. Phillips, M. et al. Breath biomarkers of active pulmonary tuberculosis. *Tuberculosis* **90**, 145–151 (2010).
12. Phillips, M. et al. Point-of-care breath test for biomarkers of active pulmonary tuberculosis. *Tuberculosis* **92**, 314–320 (2012).
13. Kolk, A. et al. Breath analysis as a potential diagnostic tool for tuberculosis. *Int. J. Tuberc. Lung Dis.* **16**, 777–782 (2012).
14. Dang, N. A., Janssen, H.-G. & Kolk, A. H. Rapid diagnosis of tb using GC-MS and chemometrics. *Bioanalysis* **5**, 3079–3097 (2013).
15. Beccaria, M. et al. Exhaled human breath analysis in active pulmonary tuberculosis diagnostics by comprehensive gas chromatography-mass spectrometry and chemometric techniques. *J. Breath Res.* **13**, 016005 (2018).
16. Mellors, T. R. et al. Identification of mycobacterium tuberculosis using volatile biomarkers in culture and exhaled breath. *J. Breath Res.* **13**, 016004 (2018).
17. Fu, L. et al. A cross-sectional study: A breathomics based pulmonary tuberculosis detection method. *BMC Infect. Dis.* **23**, 148 (2023).
18. Arnold, C. Diagnostics to take your breath away. *Nat. Biotechnol.* **40**, 990 (2022).
19. Chen, D., Bryden, W. A. & Wood, R. Detection of tuberculosis by the analysis of exhaled breath particles with high-resolution mass spectrometry. *Sci. Rep.* **10**, 7647 (2020).
20. Bobak, C. A. et al. Breath can discriminate tuberculosis from other lower respiratory illness in children. *Sci. Rep.* **11**, 2704 (2021).
21. Loots, D. T. An altered mycobacterium tuberculosis metabolome induced by KATG mutations resulting in isoniazid resistance. *Antimicrob. Agents Chemother.* **58**, 2144–2149 (2014).
22. Xue, J. et al. Single quadrupole multiple fragment ion monitoring quantitative mass spectrometry. *Anal. Chem.* **93**, 10879–10889 (2021).
23. Wenig, P. & Odermatt, J. Openchrom: A cross-platform open source software for the mass spectrometric analysis of chromatographic data. *BMC Bioinformatics* **11**, 1–9 (2010).
24. Stein, S. E. An integrated method for spectrum extraction and compound identification from gas chromatography/mass spectrometry data. *J. Am. Soc. Mass Spectrom.* **10**, 770–781 (1999).
25. Trabue, S. L., Anhalt, J. C. & Zahn, J. A. Bias of tedlar bags in the measurement of agricultural odorants. *J. Environ. Qual.* **35**, 1668–1677 (2006).
26. Worton, D. R. et al. Improved molecular level identification of organic compounds using comprehensive two-dimensional chromatography, dual ionization energies and high resolution mass spectrometry. *Analyst* **142**, 2395–2403 (2017).
27. Maitra, A. et al. Cell wall peptidoglycan in mycobacterium tuberculosis: An achilles' heel for the tb-causing pathogen. *FEMS Microbiol. Rev.* **43**, 548–575 (2019).
28. Meacci, F. et al. Drug resistance evolution of a mycobacterium tuberculosis strain from a noncompliant patient. *J. Clin. Microbiol.* **43**, 3114–3120 (2005).

Acknowledgements

The authors extend their gratitude to all the participants who generously took part in this study, including the TB patients, MDR-TB patients, and control subjects. Special thanks are also due to the Ministry of Health and Wellness in Botswana, as well as the staff and facilities of Nyangabgwe, Sekgoma, and Princess Marina Hospitals, Palapye Primary Hospital, and Palapye clinics (namely, Kediretswe Clinic, Extension 3 Clinic, and Khurumela Clinic) for their invaluable support throughout this research. This work was funded by the European Union (Grant no. DCI-PANAF/2020/420-028) under the African Research Initiative for Scientific Excellence (ARISE) pilot programme. ARISE is implemented by the African Academy of Sciences with the support of the European Commission and the African Union Commission. The authors bear sole responsibility for the content of this publication, which does not necessarily reflect the views or positions of the European Union, the African Academy of Sciences, or the African Union Commission.

Author contributions

G.C. Is the principal investigator, he conceived the experiment(s) and supervised the whole project, A,G. M. and T,C. S. Conducted the experiment(s) and analysed the results, with A,G.M. being the lead author, S.B. analysed the results. All authors reviewed the manuscript.

Declarations

Competing interests
The authors declare no competing interests.

Additional information
I George Chimowa as the corresponding author and on behalf of the other co-authors declare that we have no actual, potential or perceived competing interest in relation to publication of this work. The research work was carried out with due diligence and professional conduct.

Additional information
Correspondence and requests for materials should be addressed to G.C.

Reprints and permissions information is available at www.nature.com/reprints.

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Open Access This article is licensed under a Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 International License, which permits any non-commercial use, sharing, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if you modified the licensed material. You do not have permission under this licence to share adapted material derived from this article or parts of it. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit http://creativecommons.org/licenses/by-nc-nd/4.0/.

© The Author(s) 2025