Supplementary Information

Niche-specific metabolic signatures in non-typeable *Haemophilus influenzae*

Keshab Bhattarai[1†], Bikash Baral[2†], Aleksandra Sarnowicz[1], Margo Diricks[3,5], Stefan Niemann[3,6,7,8], Jan Rupp[4,5,6], Katarzyna A. Duda[1*]

[1]Group of Biofunctional Metabolites and Structures, Research Center Borstel, Leibniz Lung Center, Member of German Center for Lung Research (DZL), Airway Research Center North, Borstel, Germany

[2]Organismal and Evolutionary Biology Research Programme, University of Helsinki, FI-00014 Helsinki, Finland

[3]Research Group of Molecular and Experimental Mycobacteriology, Research Center Borstel, Leibniz Lung Center, Borstel, Germany

[4]Institute of Medical Microbiology, University Hospital Schleswig Holstein/ Campus Lübeck, Lübeck, Germany

[5]Infectious Disease Clinic, University Hospital Schleswig-Holstein/ Campus Lübeck, Lübeck, Germany

[6]German Center for Infection Research (DZIF), Hamburg-Lübeck-Borstel-Riems, Germany

[7]Laboratoire Biologie Intégrative des Populations, Evolution Moléculaire, EPHE, PSL University, Paris, France.

[8]Institut de Systématique, Evolution, Biodiversité, ISYEB, Muséum national d'Histoire naturelle, CNRS, Sorbonne Université, EPHE, Université des Antilles, Paris, France.

*** Correspondence:**
Corresponding Author, kduda@fz-borstel.de

[†]These authors contributed equally to this work.

1 Supplementary Information

1.1 Materials and Methods

Whole-Genome-based taxonomy (TYGS)

The genome sequences of the nine isolated strains were analyzed using the TYGS platform (https://tygs.dsmz.de) for taxonomic classification referencing updated LPSN data (https://lpsn.dsmz.de). Closest type strains were identified by (i) MASH-based genome comparisons against all TYGS type strains, and (ii) 16S rRNA gene extraction (via RNAmmer), followed by BLAST searches against 23,259 type strains to select the top 50 hits by bitscore. Genome-based distance calculations were performed using GBDP (algorithm: coverage; formula: d5), and the 10 closest type strains per genome were selected for further analysis. Pairwise GBDP comparisons (algorithm: trimming; formula: d5) supported phylogenomic inference with 100 replicates. Digital DNA-DNA hybridization (dDDH) values and confidence intervals were computed using GGDC 4.0. A minimum evolution tree with SPR postprocessing was generated using FASTME, supported by 100 pseudo-bootstrap replicates and midpoint rooting. Phylogenies were visualized using PhyD3. Species and subspecies delineation was defined by dDDH thresholds of 70% and 79%, respectively.

1.2 Results

Distinct phylogenetic lineages

To validate phylogenetic relationships at the whole-genome level, genome-wide comparisons were conducted using the Type Strain Genome Server (TYGS), employing the Genome BLAST Distance Phylogeny (GBDP) method. GBDP analysis confirmed the close genetic proximity of strains H188, H189, and H182 to the reference *Haemophilus influenzae* strain CP000057.2, supporting 16S rRNA-based findings with high-resolution genomic data (Supplementary Figure S1 and S2). Species and subspecies clustering results are shown in Supplementary Table S3, and taxonomic assignments in Supplementary Table S2. In total, 41 species clusters were defined, with the query strains distributed across four species and eight of 46 subspecies clusters. These results (Supplementary Figure S2) highlight the effectiveness of integrating 16S and whole-genome data for precise taxonomic resolution (Supplementary Table S3_Sheet 1 and 2).

Species-level taxonomic resolution was further refined through a comprehensive Average Nucleotide Identity (ANI) analysis conducted across all 9 bacterial isolates. Following established thresholds, ANI values ≥96% are indicative of conspecific relationships. The majority of isolates exhibited ANI values exceeding this threshold, thereby supporting their close genetic affiliation with members of the *Haemophilus* genus (Supplementary Table S1_Sheet 4).

Conserved metabolic capacity with limited strain-specific variation

A comprehensive genomic analysis of carbohydrate metabolism pathways was conducted across the nine Haemophilus strains sourced from the lungs (n = 5), ear smear (n = 2), and pharyngeal smear (n = 2) as well as for the reference H. influenzae CP000057.2 (H. influenzae 86-028NP) genome. Annotation using KAAS and ortholog profiling revealed a broadly conserved core set of metabolic genes underpinning central carbon metabolism, alongside targeted, niche-correlated gene losses indicative of differential metabolic association (Fig. S3, Supplementary Table S4_Sheets 1, 2, and 3).

Among 19 targeted metabolic genes, 17 were conserved across the strains, including pfkA, pyk, pckA, and the fructose-1,6-bisphosphatase isoforms (fbp, glpX). The only exception was frmA, which was absent in the reference strain. Notably, the ear strains (H104, H213) retained the complete set. While this may suggest metabolic robustness within this niche, the limited number of ear strains warrants caution, and broader sampling would be required to confirm whether this pattern reflects niche-specific metabolic stability or simply isolate-level variation. Pyruvate metabolism (26 genes) showed near-complete conservation, with frmA again missing in the reference strain. Likewise, starch and sucrose metabolism genes (glgA, glgB, galU, pgi, glgC, malQ, crr) were conserved, although glycogen phosphorylase (glgP) was universally absent, suggesting divergence in carbohydrate storage as compared to the reference strain. The Leloir pathway (galK, galT, galM, galU/UGP2) was also intact, but galE was consistently undetected, possibly reflecting annotation gaps or alternative mechanisms of galactose-1-phosphate conversion into UDP-galactose (Supplementary Table S4_Sheets 1, 2 and 3).

In contrast, the tricarboxylic acid (TCA) cycle and related pathways showed localized variation. Of 13 TCA genes, 12 were conserved, including mdh, aceE, pdhC, lpd, and frdA-D. The β-subunit sucC was absent only in lung strain H182, while its α-subunit (sucD) and other components remained intact. Similarly, propanoate and butanoate pathways were broadly conserved, with ackA, pta, accA-D, mgsA, frdA-D, atoA, ilvH, and ilvN present. However, sucC was again missing in H182, pta in H149, and pflD was universally absent. Large acetolactate synthase subunits (ilvB, ilvG, ilvI) appeared sporadically, mainly in PS strain H149. Glyoxylate and dicarboxylate metabolism was largely conserved (mdh, glyA, ACAT, glxK, gph, purU, eda, glnA), though lung strains showed selective absences (ACAT, glxK, eda in H117; glxK in H212). By contrast, inositol phosphate metabolism was fully conserved across all strains.

Sugar interconversion and accessory pathways exhibited greater variability. Within the pentose phosphate pathway (PPP), 14 of 17 genes were conserved, but kdgK and eda were absent in lung strain H212, while fbp was missing in pharyngeal strain H189. Pentose and glucuronate interconversion genes consistently included galU, RPE, fucK and fucI, whereas other genes like xylA, xylB, fucA, uxuA and uxaC were selectively absent, particularly in H212 (Supplementary Table S4_sheet 2 and 3).

Fructose and mannose metabolism was mostly conserved (fbaA, tpiA, fruA, fbp), though fucA genes were absent in H213, and annotation gaps were noted for fruK and manB. Finally, C5-branched dibasic acid metabolism was largely conserved across all strains, with key enzymes such as leuB, leuC, leuD, ilvH, ilvN, and sucD present in every genome examined. However, notable strain-specific variation was observed in the acetolactate synthase large subunits (ilvB, ilvG, ilvI), which were exclusively detected in strain H149 (Supplementary Table S4_sheets 2 and 3). This suggests that while the core pathway remains broadly intact, the capacity for certain branched-chain amino acid biosynthetic reactions may be uniquely maintained in H149. Collectively, these results highlight the strong conservation of central metabolism, alongside selective gene losses indicative of niche-specific remodelling and functional redundancy.

Conserved metabolic genes despite heterogeneous metabolomes

Within purine metabolism, three genes deoD (K03784), LACC1 (K05810), and ushA (K11751) were associated with inosine interconversion to hypoxanthine, inosine synthesis from adenosine, and inosine synthesis from inosine monophosphate (IMP), respectively. Additionally, we detected two

genes directly involved with hypoxanthine: hprT (K00760), which catalyzes hypoxanthine to IMP, and again deoD (K03784), which catalyzes the interconversion of hypoxanthine and deoxyinosine. This indicates that all nine strains showed comparable genomic potential for inosine and hypoxanthine metabolism despite differences in their extracellular metabolite abundances.

In nicotinate and nicotinamide metabolism, cobB (K12410) was identified as directly involved in NAD+ conversion to niacinamide, while nadR (K06211) and deoD (K03784) provided an alternative route through intermediate steps.

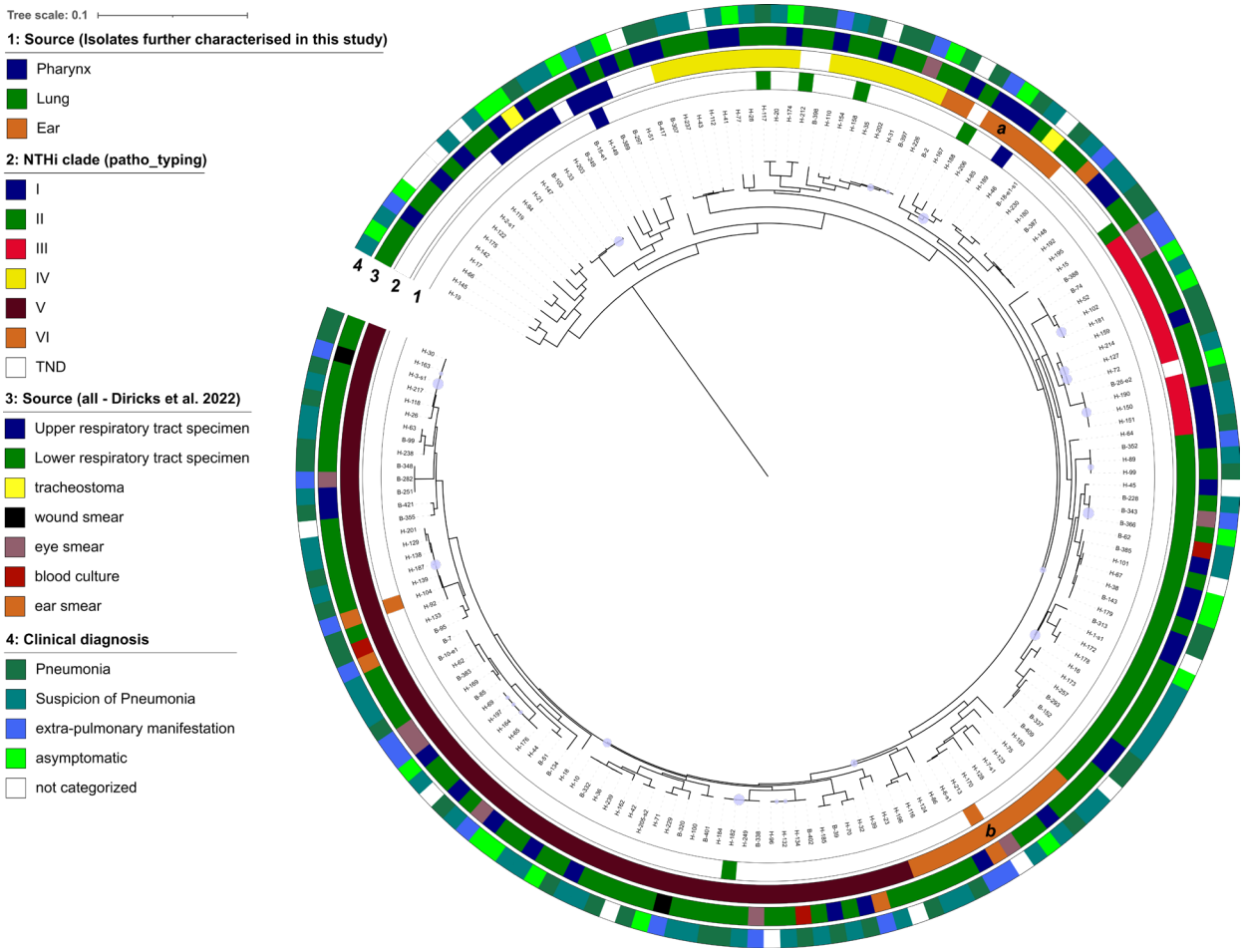
In pyrimidine metabolism, udp (K00757) was associated with uridine formation from uridine and deoxyuridine. The phenylalanine metabolism genes aspC (K00813) and hisC (K00817), both encoding aminotransferases capable of converting L-phenylalanine to phenylpyruvate, were present in all nine strains.

In valine, leucine and isoleucine degradation pathway, we detected a gene, E2.6.1.42 (K00826), that catalyzes the L-isoleucine interconversion to (S)-3-Methyl-2-oxopentanoic acid and L-valine interconversion to 3-methyl-2-oxobutanoate. This gene was present in all 9 strains. We also detected a complete set of genes, E2.2.1.6L (K01652), E2.2.1.6S (K01653), ilvC (K00053), ilvD (K01687), E2.6.1.42 (K00826), and alaA (K14260), that catalyze pyruvate to L-valine and L-isoleucine through multiple intermediate steps.

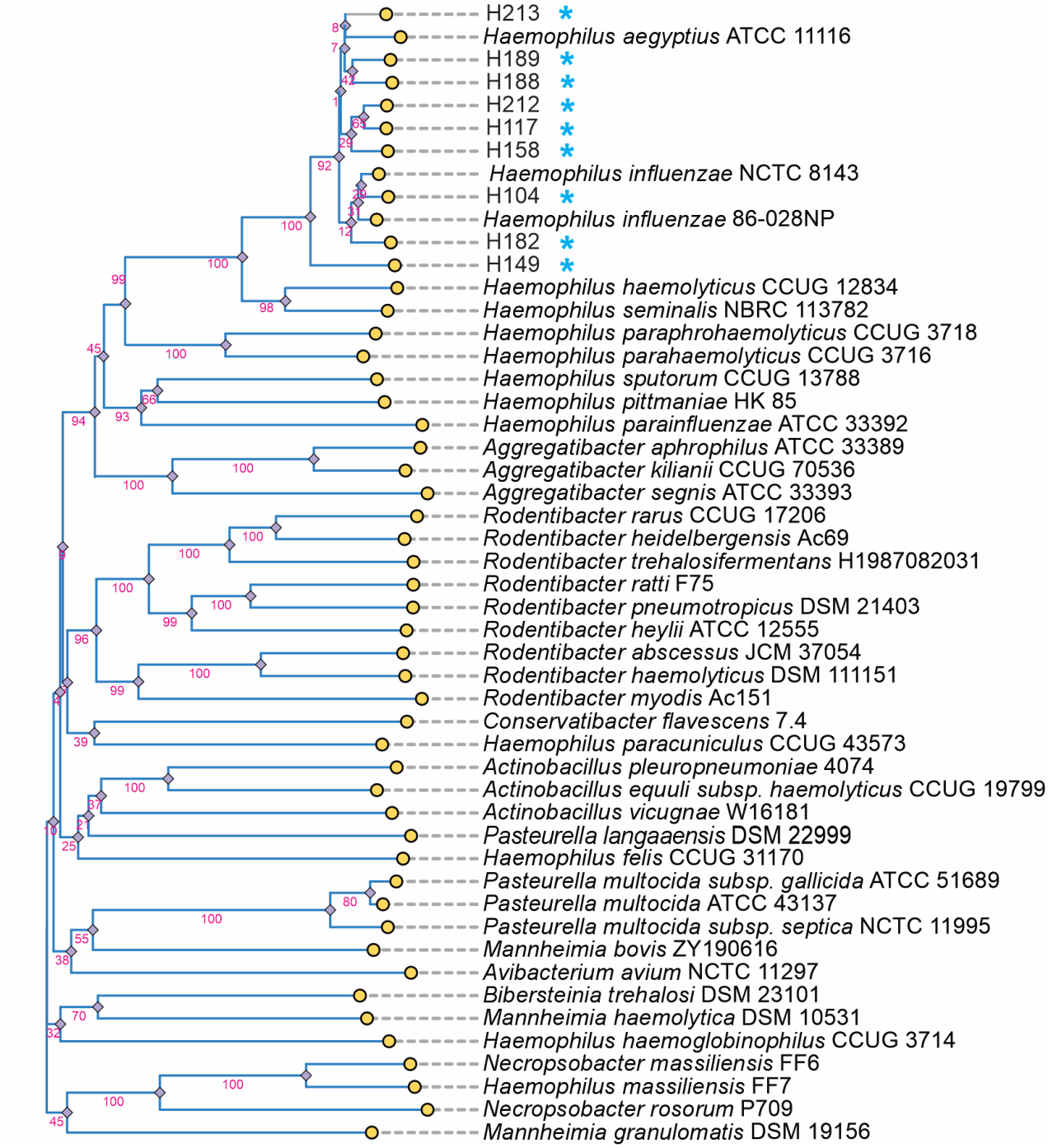
In vitamin B6 metabolism, we detected a gene pdxK (K00868) that catalyzes pyridoxine to pyridoxine-5’ phosphate and another gene pdxH (K00275) that catalyzes the interconversion of pyridoxine to pyridoxal and further to pyridoxamine.

Despite this overall conservation, a small number of site-specific KO differences were observed. In purine metabolism, K20881 (yrfG, IMP to Inosine) was unique to ES, K25589 (inuH, inosine to hypoxanthine, uridine to uracil) was unique to Lung, and K01081 (E3-1-3-5, IMP to Inosine, AMP to adenosine, UMP to uridine) was unique to ES and Lung, while the rest were present in all groups. In nicotinate and nicotinamide metabolism, K01081 was unique to ES and Lung. In pyrimidine metabolism, K25589 was restricted to Lung, while K01081 was unique to ES and Lung

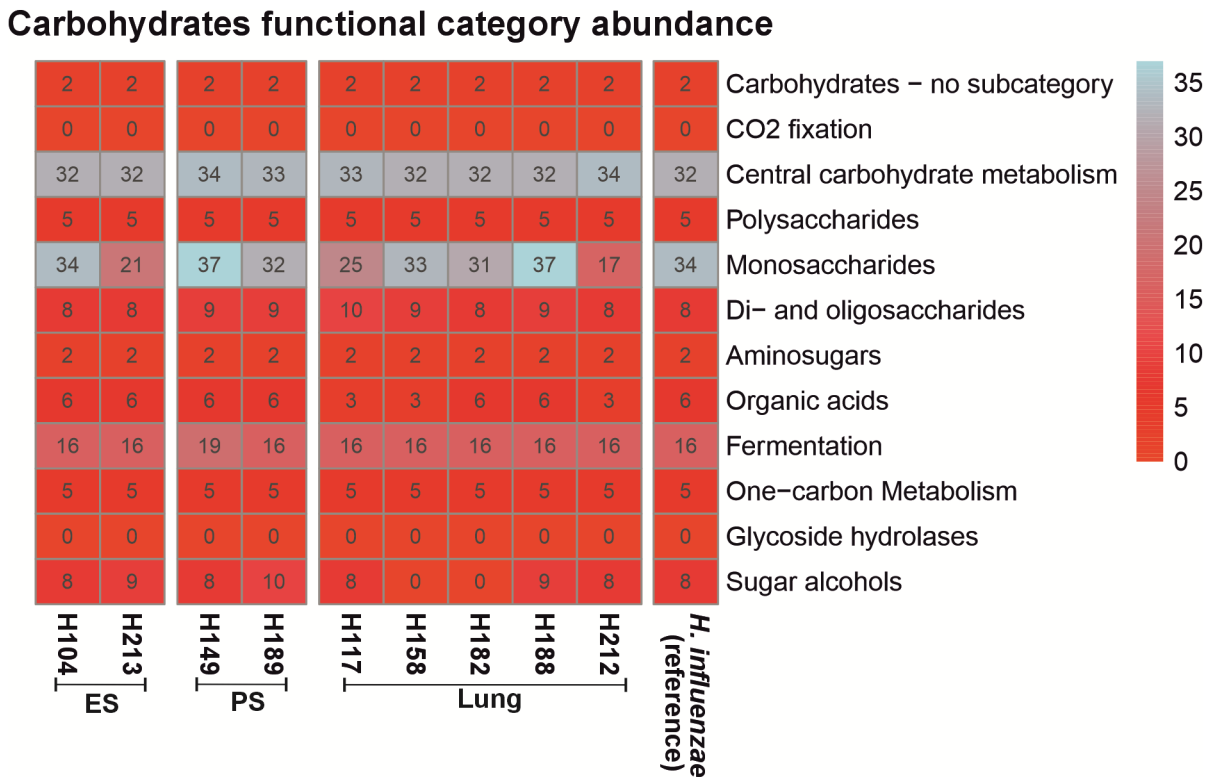
1.3 Supplementary Figures



Supplementary Fig. S1. Whole genome sequence-based phylogenetic tree of 172 H. influenzae isolates derived from Diricks et al.(2022)(Diricks et al., 2022). The approximately maximum-likelihood tree was previously constructed based on 104 core genes using FastTree. Local bootstrap values below 0.9 are highlighted with blue circles. NTHi clades were previously determined using the bioinformatic tool patho_typing(Pinto et al., 2019) based on gene absence/presence patterns. For isolates that were not classified by patho typing in one out of six previously defined clades (TND), clade affiliation was inferred from phylogenetic positioning. Clade VI was further divided into subgroups a and b based on their positions within the current tree.

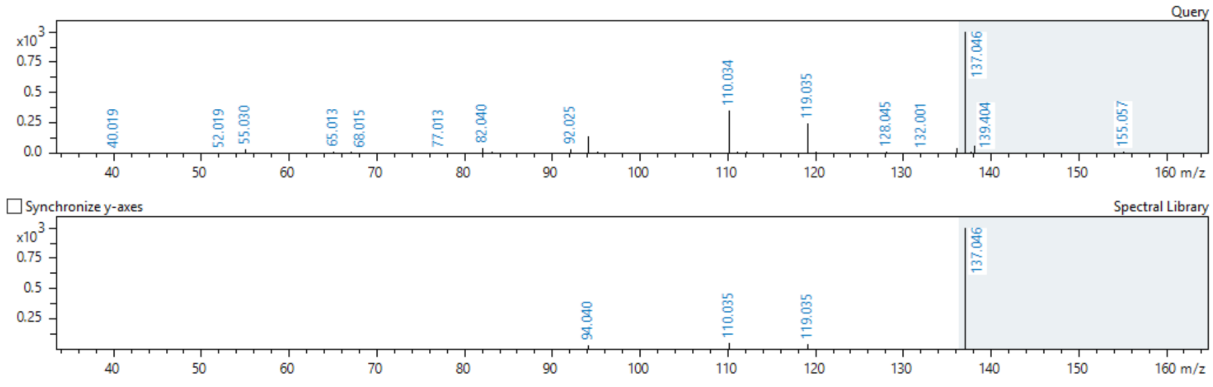


Supplementary Fig. S2. GBDP distance-based phylogenetic tree obtained from 16S rRNA gene sequences generated using FastME 2.1.6.1. Using the GBDP distance formula d5, branch lengths are proportionate to distances. Bootstrap support values >60% from 100 pseudo-replicates are indicated above branches, with a mean branch support of 63.0%. The tree was rooted at the midpoint. Species with blue asterisks (*) are the strains currently studied.

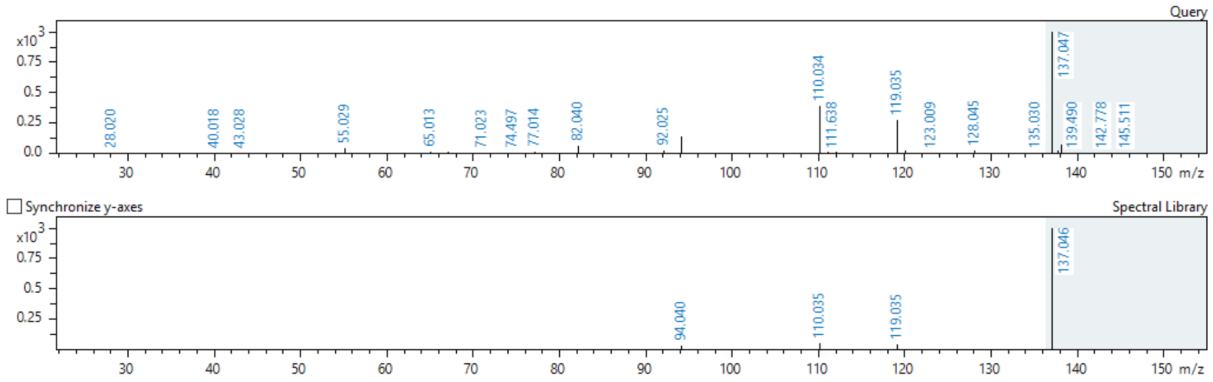


Supplementary Fig. S3. Carbohydrate metabolism encoding genes found in different studied Haemophilus influenzae strains. The numbers represent the number of KO (KEGG ORTHOLOGY) entries retrieved by KAAS from the KEGG database.

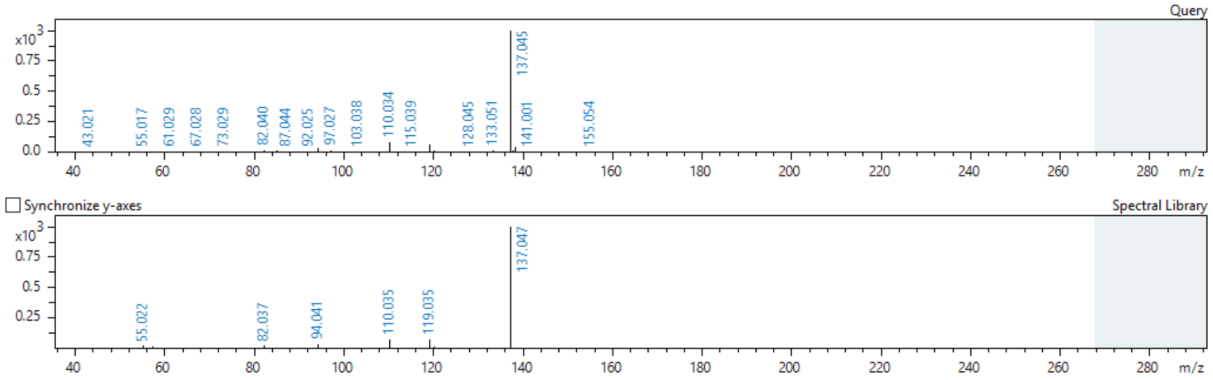
1.2_137.04562_Hypoxanthine (MS^2 score: 971)_MSMS Public



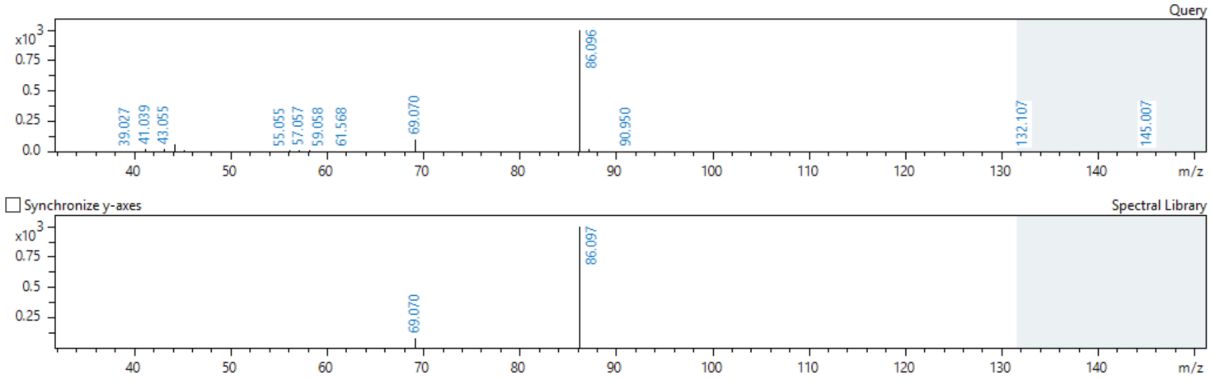
1.76_137.04561_Hypoxanthine (MS^2 score: 969)_MSMS Public



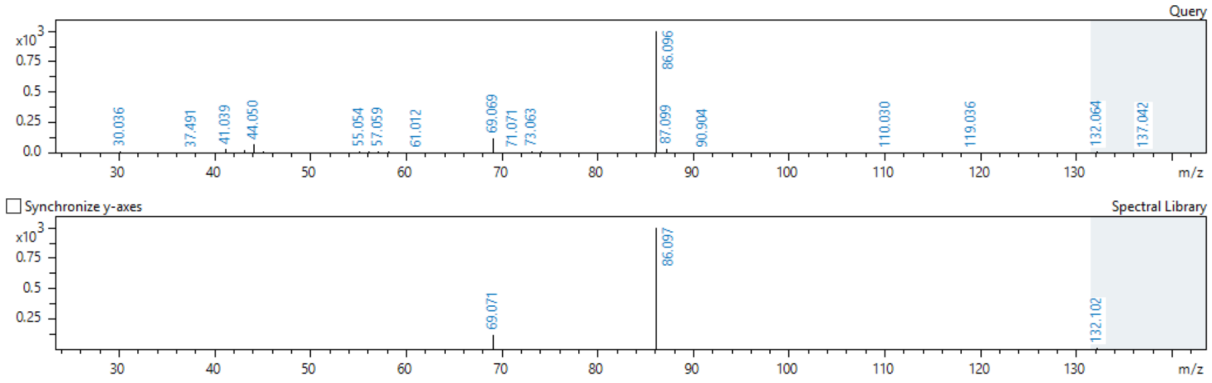
1.2_269.0885_Inosine (MS^2 score: 998)_MSMS Public



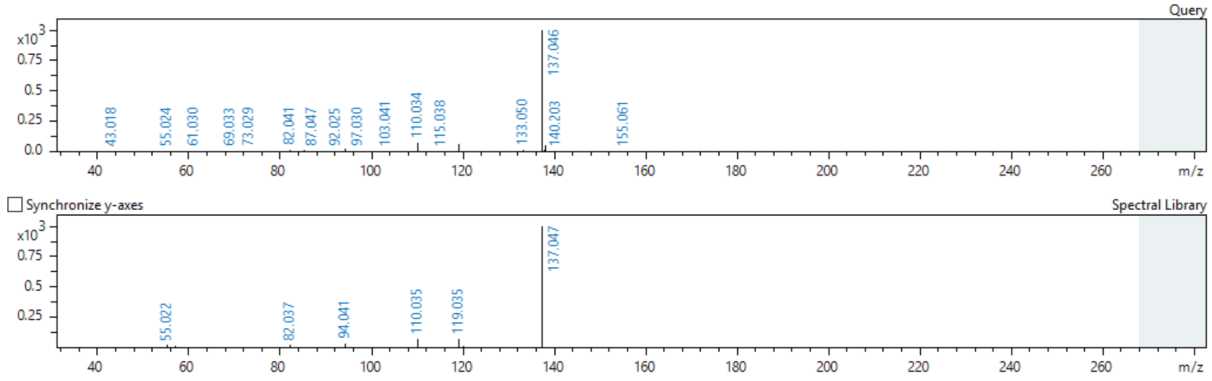
1.77_132.10193_Isoleucine (MS^2 score: 997)_MSMS Public



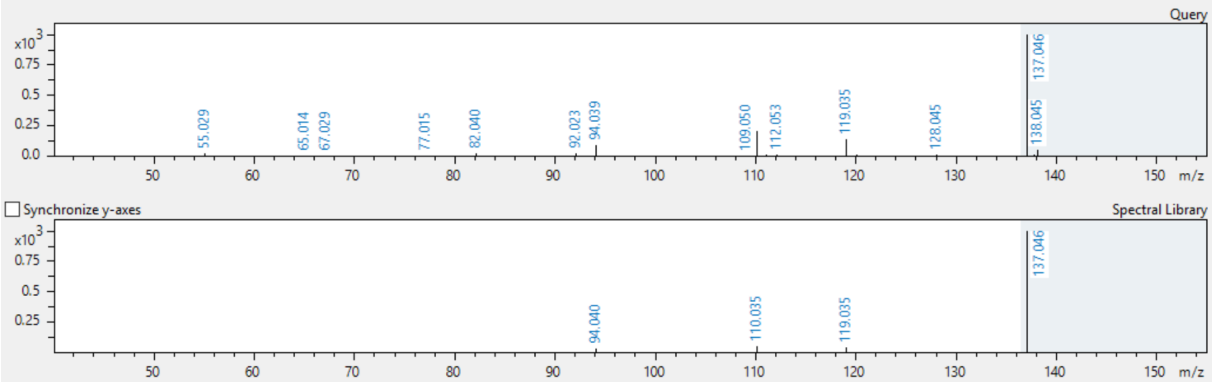
1.2_132.10187_Isoleucine (MS^2 score: 996)_MSMS Public



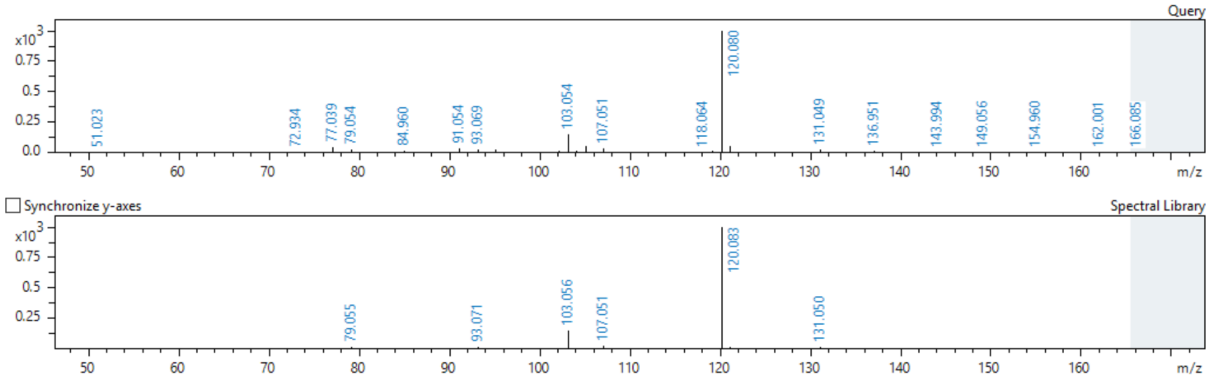
1.77_269.08863_Inosine (MS^2 score: 998)_MSMS Public



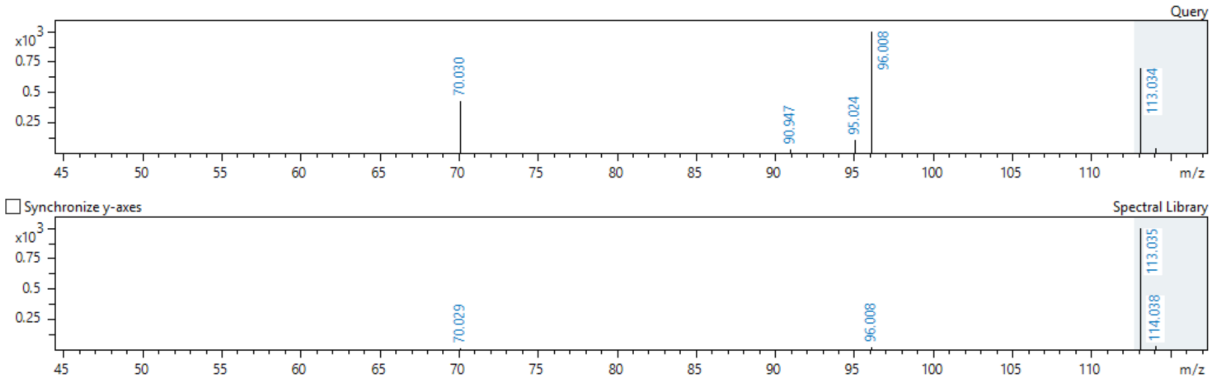
1.61_137.04571_Hypoxanthine (MS^2 score: 975)_MSMS Public



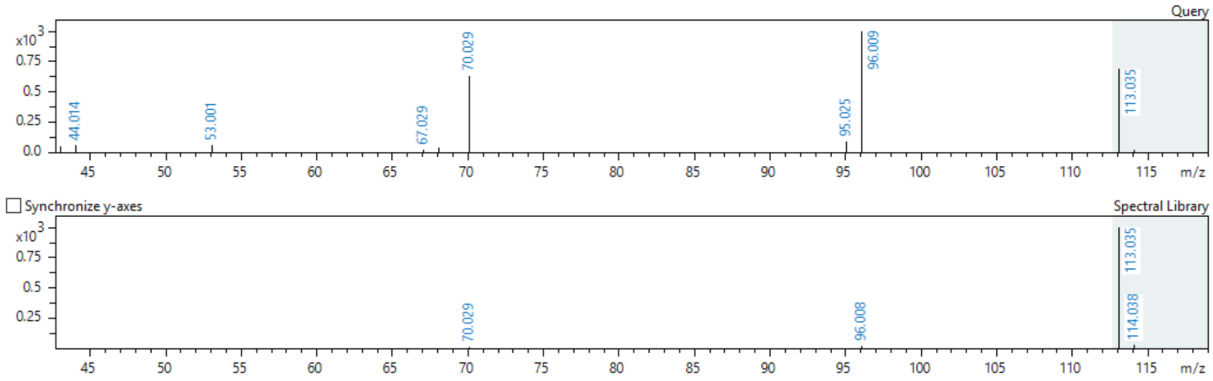
2.24_166.08643_Phenylalanine (MS^2 score: 997)_MoNA



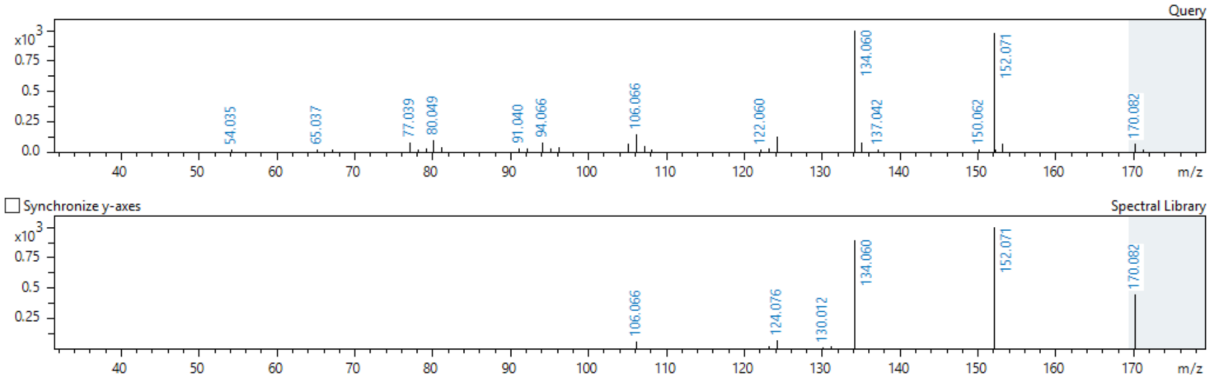
1.58_113.03432_Uracil (MS^2 score: 991)_MSMS Public



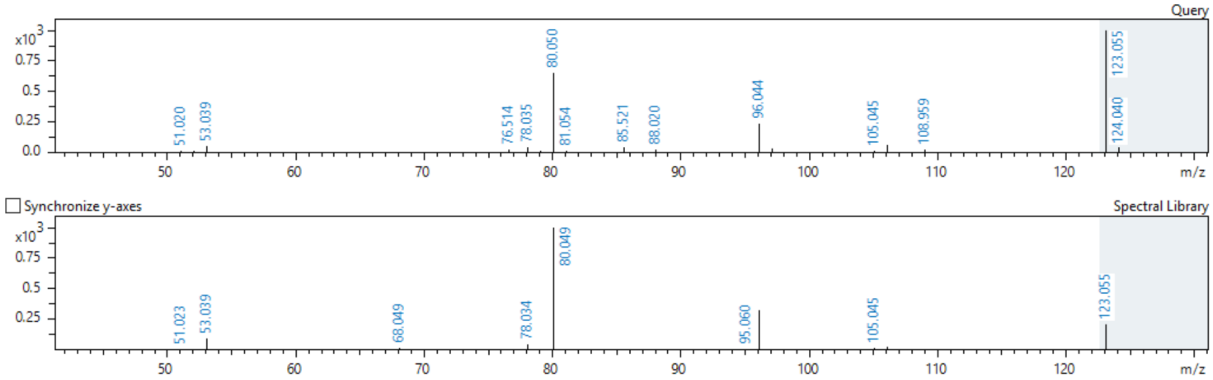
1.22_113.03432_Uracil (MS^2 score: 988)_MSMS Public



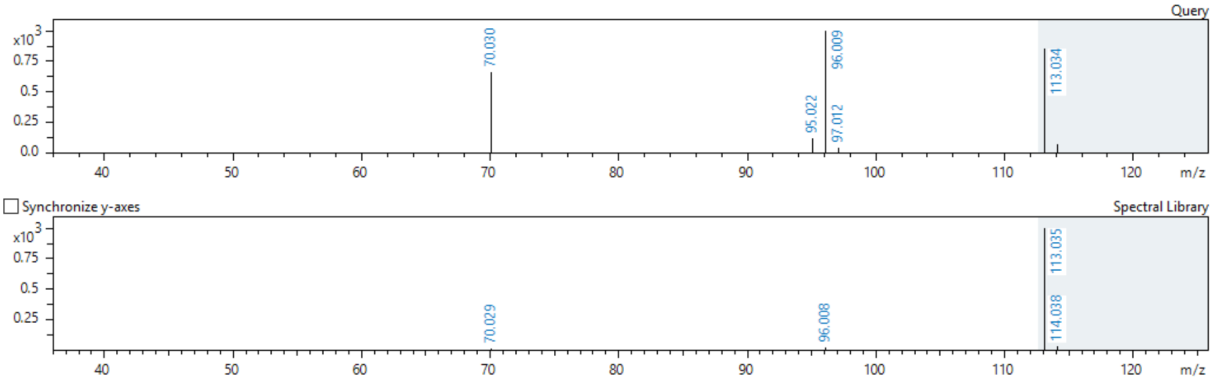
1.17_170.08153_Pyridoxin (MS^2 score: 983)_MSMS Public



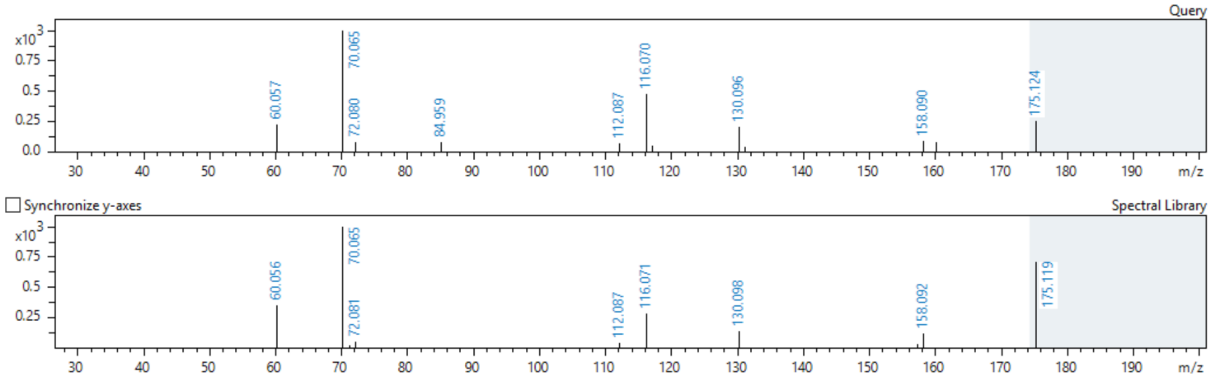
1.76_123.0551_Niacinamide (MS^2 score: 993)_MoNA



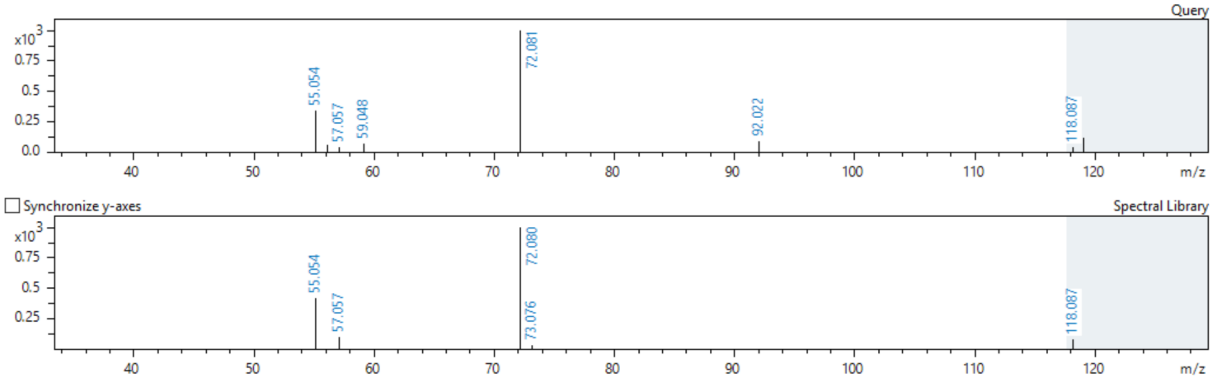
1.75_113.03433_Uracil (MS^2 score: 988)_MSMS Public



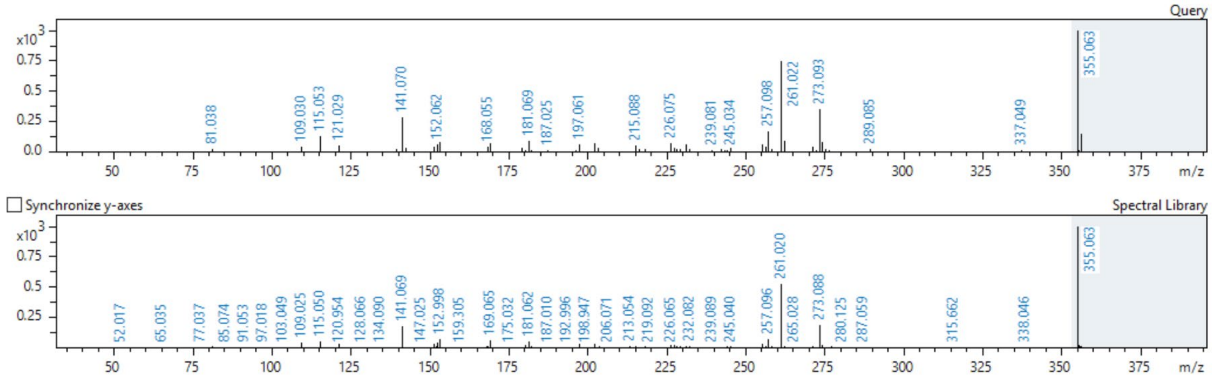
1.06_175.11932_Arginine (MS^2 score: 970)_MSMS Public



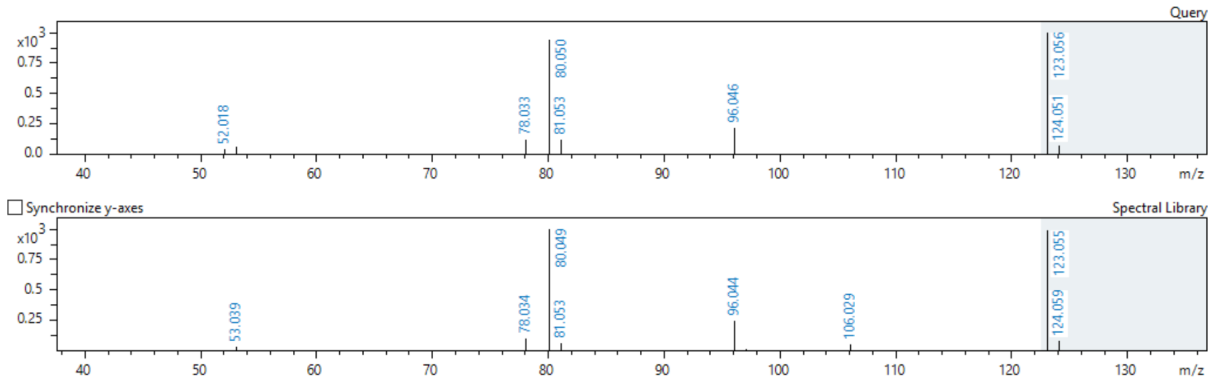
1.17_118.08616_Valine (MS^2 Score: 990)_MSMS Public



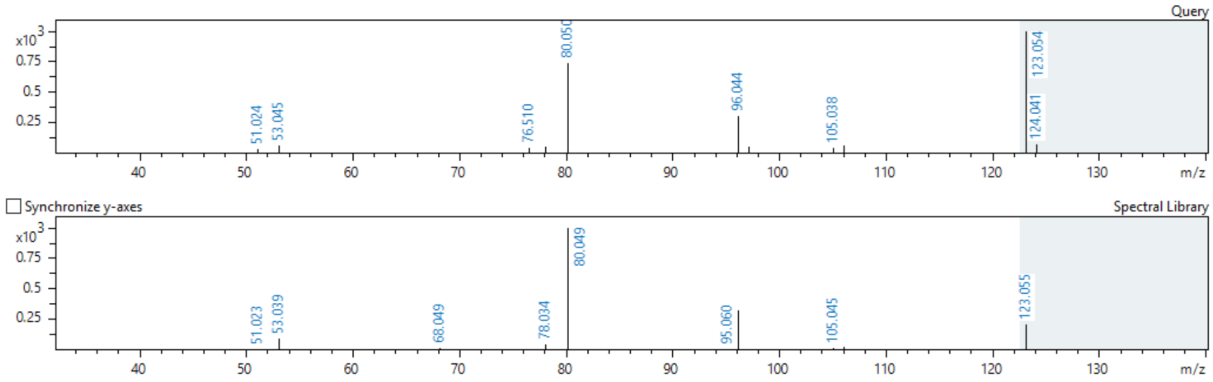
6.97_355.06363_Phenolsulfonphthalein (MS^2 score: 976)_Bruker MetaboBASE 3.0



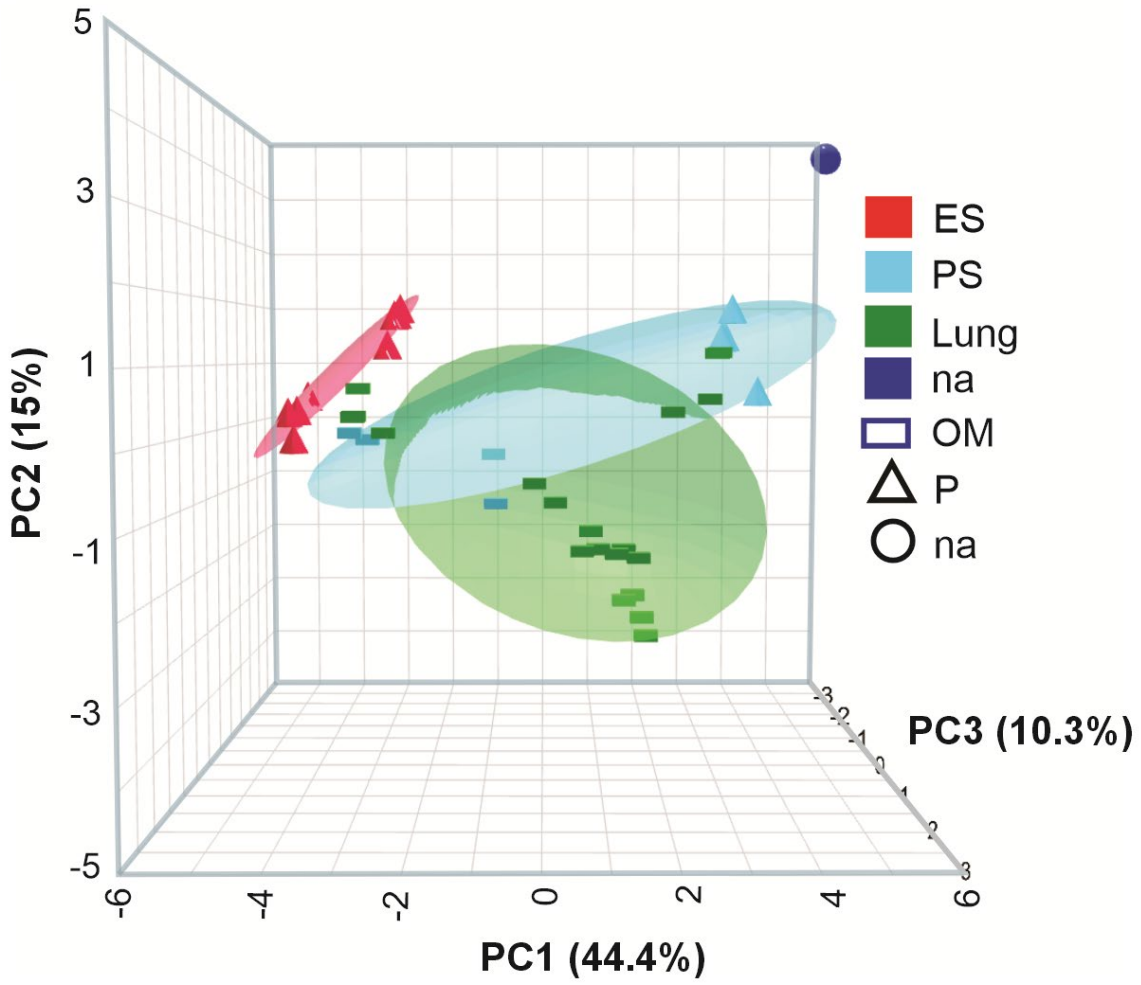
1.2_123.05522_Niacinamide (MS^2 score: 995)_MSMS Public



1.52_123.05523_Niacinamide (MS^2 score: 992)_MoNA



Supplementary Fig. S4. MS^2 spectral match of the metabolites from 9 NTHi isolates against MS^2-libraries. (Labelling: retention time_*m/z*_compound name (MS^2 score)_library name)



Supplementary Fig. S5. 3D PCA plot of site-specific samples based on metabolites intensities with metadata overlay. (ES = ear sample, PS = pharyngeal sample, na = media control, OM = otitis media and P = pneumonia).

1.4 Captions of Supplementary Tables

Suppelementary Table S1. Genomic assembly quality, structural features, and functional landscape of *Haemophilus influenzae* strains.

Supplementary Table S2. Comparative genomic visualization of NTHi strains generated using Proksee.

Supplementary Table S3. Genome-wide phylogenomic analysis of *H. influenzae* isolates based on GBDP clustering.

Supplementary Table S4. Distribution of KEGG orthologs (KOs) and corresponding gene annotations linked to carbohydrate metabolism. The listed KOs and gene annotations encompass major carbohydrate utilization and biosynthetic pathways across *Haemophilus* strains. The table highlights strain-specific presence/absence patterns reflecting metabolic conservation and diversity within the carbohydrate metabolic network.

Supplementary Table S5. Biosynthetic gene cluster (BGC) regions of different NTHi strains analysed by antiSMASH version 8.0.2 from the annotated whole genome sequence files (.gbk file format).

Supplementary Table S6. List of 38 metabolites detected using LC-MS and their signal intensities in the respective samples.

Supplementary Table S7. Significant metabolites with P-cutoff: 0.05 (FDR) varied in different groups without covariate adjustment. The compared groups are ES vs Lung (a), ES vs PS (b) and Lung vs PS (c).

Supplementary Table S8. Annotated KEGG orthology groups for different pathways present in individual bacterial isolates.

2 References

Diricks, M., Kohl, T. A., Käding, N., Leshchinskiy, V., Hauswaldt, S., Jiménez Vázquez, O., et al. (2022). Whole genome sequencing-based classification of human-related *Haemophilus* species and detection of antimicrobial resistance genes. *Genome Med.* 14, 13. doi: 10.1186/s13073-022-01017-x

Pinto, M., González-Díaz, A., Machado, M. P., Duarte, S., Vieira, L., Carriço, J. A., et al. (2019). Insights into the population structure and pan-genome of *Haemophilus influenzae*. *Infect. Genet. Evol.* 67, 126–135. doi: 10.1016/j.meegid.2018.10.025