Gas Chromatography-Mass Spectrometry Data used for "Identifying Palaeolithic birch tar production techniques: challenges from an experimental biomolecular approach"

This Dataset is accompanying the article:

Kozowyk, P.R.B., Baron, L.I. & Langejans, G.H.J. Identifying Palaeolithic birch tar production techniques: challenges from an experimental biomolecular approach. Sci Rep 13, 14727 (2023). https://doi.org/10.1038/s41598-023-41898-5.

Introduction:

The dataset contains data collected between September 2019 and September 2022 as part of the Ancient Adhesives project under the European Union's Horizon 2020 research and innovation programme Grant Agreement No. 678 804151 (Grant holder G.H.J.L.). It is being made public to act as supplementary data for a publication and for other researchers to use this data in their own work. Due to the nature of the GC-MS data, all raw files have been supplied which will allow further processing in the original software.

The dataset includes 8 zip files with the raw GC-MS data. Each zip file contains the files necessary to open and manipulate the data using the original software Agilent OpenLab 2.5; 1 xlsx file containing processed information used in the manuscript; 1 zip file containing CSV files used for principal component analysis in the manuscript. Methodological data for the collecting and generating the data can be found in the accompanying manuscript in Scientific Reports https://doi.org/10.1038/s41598-023-41898-5.

Zip files containing raw GC-MS data. Files included are DX (for opening with Agilent OpenLab 2.5) and accompanying ACAML, DX, MFX, BIN, RX, PMX, and AMX files:

210927 Hor_AM3.zip

210927 Hor_PR11.zip

210927 Hor_RS7.zip

211007 Hor_3201.1.zip

211007 Vla_PR11.2.zip

211007 Vla_RS02.2.zip

211014 Mct_3.1.zip

211014 Mct_4.1.zip

Xlsx file containing processed information used in the accompanying publication:

230118 Supplementary Information.xlsx

This file contains 12 Sheets:

Sheet 1:

A summary of all the peaks automatically identified using Agilent Openlab and the NIST (National Institute of Standards and Technology) library.

Row 1 contains File name (corresponding to the raw GC-MS files)

Row 2 contains ID used in the manuscript

Row 3 contains GCMS peak information:

RT = Retention time (min)

Compound name = compound automatically identified from NIST library using Agilent Openlab corresponding to the RT.

Score = Match factor, or a comparison of the unknown's mass spectrum's peak to those of the peaks in the library's spectra.

Rev. Score = The match factor when the peaks in the unknown's spectrum that are not in the library's known reference spectrum are ignored.

Prob.% = Probability percent that the compound is identified correctly within the NIST library.

CAS # = Chemical Abstracts Service Number – A unique accession number assigned to a given compound

Library Id = Identification number from the NIST mass spectral library

Sheet 2:

A summary of the presence/absence of specific peaks used in the accompanying manuscript.

DP/SP = double pot and single pot methods of tar production.

Sheets 3-12:

Complete GC-MS data exported for samples analyzed in the accompanying publication.

RT= Retention time (min)

Type = Type of integration. BB = baseline to baseline, BV = baseline to valley, VB = valley to baseline, VV = valley to valley, MV = manual to valley, VM = valley to manual, MB = manual to baseline, MM = manually integrated

Width (min) = Peak width

Area = Peak area

Height = Peak height

Area % = Peak area %

Zip file containing CSV files generated for the Principle Component Analysis conducted in the accompanying publication. Relevant data on how the analysis was conducted can be found in the manuscript.

clustvisPCAloadings

clustvisPCAscores

ClustvisPCAcartable

clustvisProcessedDate