Dataset on the solvent screening and extraction optimization of volatile compounds from *Sequoia sempervirens* leaves

Loïc Latorre[1,2*], **Bertrand Charrier**[1]

[1]IPREM (UMR 5254 CNRS–UPPA), Université de Pau et des Pays de l'Adour, Pau, France

[2]Ministère de l'Éducation Nationale, Professeur Agrégé de Sciences Physiques option Chimie, France

*Corresponding author: llatorre@univ-pau.fr

June 25, 2026

Abstract

This dataset reports the comprehensive results of a solvent screening and extraction optimization experiment targeting volatile compounds obtained from *Sequoia sempervirens* (coast redwood) leaves. Leaf samples were collected from a single tree and branch under rigorous and identical conditions to minimize inter-individual variability. Extractions were systematically performed in duplicate using various pure and binary solvent systems. Parameters such as solvent composition, solid-to-liquid ratio, and maceration duration were varied to map extraction efficiency. Each extract was analyzed by Gas Chromatography–Mass Spectrometry (GC–MS) with an optimized oven temperature program, and the signal intensities (Total Ion Current peak areas) of the identified compounds were normalized and averaged across independent replicates. This dataset provides a robust quantitative basis for optimizing solvent choice and extraction conditions within the framework of green chemistry, bio-sourced applications, and forest biorefinery development. The complete dataset is publicly available repository via Zenodo at https://doi.org/10.5281/zenodo.20836209.

Keywords: *Sequoia sempervirens*, solvent screening, GC–MS, volatile compounds, forest biorefinery, green extraction.

1 Introduction

The valorization of forest biomass is a central pillar of modern bioeconomy and sustainable development strategies. *Sequoia sempervirens* (coast redwood), an introduced conifer species in Western Europe, displays remarkable growth potential and resilience under temperate Atlantic conditions. Its foliage represents an abundant but currently underutilized byproduct of forestry management.

Redwood leaves contain a rich and diverse array of secondary metabolites—predominantly monoterpenes, sesquiterpenes, and diterpenes—which possess high-value potential as natural biopesticides, bioenergy additives, or biochemical building blocks [4, 5]. However, optimizing the extraction of these volatile profiles requires a systematic understanding of solvent-matrix interactions. This data paper describes a structured dataset resulting from a comprehensive solvent screening designed to map the extraction efficiency and selectivity of single and binary solvent mixtures on *Sequoia sempervirens* volatile components under distinct operating conditions.

2 Materials and Methods

2.1 Plant material and sampling

Leaves of *Sequoia sempervirens* were collected in January 2023 from a single adult tree located in southwestern France (44.545408°N, -0.498846°W). Following established sampling methodologies that demonstrate strong within-tree spatial fluctuations and geographical variations in conifer foliar monoterpenes [2, 3], the samples were taken from a single branch to minimize intra-individual variability. The leaves were separated from the twigs, then ground using a domestic mixer until obtaining fragments of 2–5 mm. The material was stored at 4 °C and processed within 24 h.

2.2 Extraction procedure

For each solvent condition, 3.00 g of ground leaves were macerated under controlled conditions at room temperature (22 ± 2 °C) using pure or binary solvent systems (acetone, ethanol, ethyl acetate, cyclohexane, and mixtures) known for their selective affinity with non-polar and semi-polar secondary metabolites [6]. The solid/liquid ratio was 1:10 (w/v) or 1:5 (w/v), and extractions were performed for 24 h or 48 h.

To establish a rigorous energy-versus-yield baseline, two distinct mass transfer modes were implemented during these macerations:

1. **Static Maceration (Simple Diffusion):** The vast majority of the experimental runs were carried out under completely static conditions without any mechanical agitation. This allowed for the characterization of passive, non-disruptive molecular diffusion from the broken leaf tissue into the surrounding liquid phase.

2. **Agitated Maceration (Forced Convection Control):** In a targeted subset of conditions, extraction vessels were subjected to continuous, controlled mechanical agitation using magnetic stirrers. This specific subset serves as a kinetic control to directly quantify the increase in mass transfer coefficients and evaluate how boundary layer turbulence affects the recovery rate of specific volatile families.

After maceration, extracts were gravity-filtered through Whatman filter paper. The solvent was evaporated under reduced pressure using a rotary evaporator (bath temperature: 30 °C; condenser temperature: 5 °C; pressure: 150 mbar). The dry extract was re-dissolved in 3.00 mL of anhydrous methanol, then divided into two GC vials for duplicate analyses.

2.3 GC–MS analysis

Analyses were carried out on a Perkin Elmer Clarus 590 gas chromatograph coupled with a Clarus SQ8S mass spectrometer, equipped with an electron impact (EI) source (70 eV), a method long validated for tracking chemical shifts and profiles in coast redwood volatile terpenes [1, 5].

Operating conditions:

- **Injector:** volume = 1 μL; carrier gas = helium; injection flow = 10 mL $\cdot$ min^{-1}; injector temperature = 300 °C.
- **Column:** Elite-5MS capillary column (30 m × 0.25 mm × 0.25 μm film thickness).
- **Oven program:** 50 °C → 190 °C at 8 °C $\cdot$ min^{-1}, then 190 °C → 300 °C at 15 °C $\cdot$ min^{-1}, hold 10 min at 300 °C.
- **Mass spectrometer:** transfer line temperature = 200 °C; ion source = 180 °C.

Identification of compounds was based on comparison of their mass spectra with the NIST 17 spectral library and retention indices, verified against biochemical database guidelines for typical conifer skeletons [7]. Peak areas were integrated from total ion chromatograms (TIC) and expressed in arbitrary units (a.u.). Each reported value corresponds to the mean of two independent extractions.

3 Dataset Description

The final dataset is structured as a clean, single tabular file (`extraction_solvent_screening_dataset_final.cs` using a tab-separator (`\t`) and UTF-8 encoding to prevent formatting conflicts with numerical values. It comprises 21 variables and 653 records.

3.1 Data Dictionary

Each row represents a specific compound detected under a unique set of extraction parameters. The columns are defined as follows:

1. `species`: Plant species (*Sequoia sempervirens*).
2. `organ`: Target plant tissue (*leaf*).
3. `solvent_mixture_name`: Unified descriptor of the solvent system, solid-to-liquid ratio, and timeline.
4. `solvent_component_1`: Primary solvent name.
5. `ratio_component_1`: Volumetric percentage of component 1 (%).
6. `solvent_component_2`: Secondary solvent name (`NA` if pure solvent).
7. `ratio_component_2`: Volumetric percentage of component 2 (%).
8. `ratio_type`: Type of ratio expression (`v/v`).
9. `solvent_ratio`: Solid-to-liquid mass/volume ratio (`1:10` or `1:5`).
10. `maceration_duration_h`: Extraction duration expressed in hours (`24` or `48`).
11. `cyclohexane_wash`: Binary indicator (`Yes`/`No`) of the cyclohexane post-treatment step.
12. `fraction_analyzed`: Target extract fraction (`primary_extract`).
13. `retention_time_min`: Gas chromatography retention time in minutes.
14. `compound_name`: Standardized international name of the detected molecule (e.g., alpha-pinene, thujene, sabinene).
15. `CAS`: Chemical Abstracts Service registry number.
16. `signal_intensity`: Average chromatographic peak area (TIC) from independent duplicates.
17. `unit_signal`: Unit of measurement (`a.u.`).
18. `analysis_method`: Analytical instrumentation used (`GC-MS`).
19. `date_extraction`: Calendar date of execution (YYYY-MM-DD).
20. `operator`: Name or initials of the laboratory operator.
21. `remark`: Supplemental qualitative observations or technical notices.

4 Technical Validation and Data Quality

To ensure the highest standard of data reliability and accuracy, several validation mechanisms were implemented:

- **Replication:** Every data point reported in `signal_intensity` represents the arithmetic mean of two completely independent extractions and subsequent chromatographic injections.
- **Precision:** The relative standard deviation (RSD) or relative gap between duplicates remained strictly below **5%** for all major volatile peaks, confirming excellent methodology repeatability.
- **Nomenclature Curation:** All compound names and CAS numbers were cross-referenced against international chemical databases to eliminate ambiguity and facilitate automated parsing by subsequent meta-analyses or algorithmic pipelines. The biocatalytic logic of target structures (such as monoterpene synthases cyclization pathways leading to pinene, thujene, or sabinene skeletons) was systematically verified against fundamental biogenic literature [7].

5 Reuse Potential and Bioeconomic Applications

This open-access dataset is intentionally tailored to support diverse fields of study:

- **Green Extraction Optimization:** Direct benchmarking of alternative, bio-sourced, or binary solvent formulations against traditional petroleum-based systems.
- **Solubility Modeling:** Providing empirical training and validation points for thermodynamic frameworks like COSMO-RS or Hansen Solubility Parameters (HSP) to model terpenoid dissolution behavior under varying operating parameters (concentration and kinetics).
- **Machine Learning & AI:** Serving as a structured, clean dataset to train predictive models for green process design in wood chemistry and sustainable forestry product development.

6 Data Availability

The structured tabular dataset file `extraction_solvent_screening_dataset_structured.csv` is hosted on the Zenodo repository and can be accessed openly via the following permanent Digital Object Identifier (DOI): https://doi.org/10.5281/zenodo.20836209.

Acknowledgements

The authors sincerely thank Charles-Henri Lannux for providing authorized access to the plant material, as well as the CNPF for their technical expertise and analytical support during the initial stages of this project.

References

[1] Okamoto, R. A., Ellison, B. O., & Kepner, R. E. (1981). Volatile terpenes in *Sequoia sempervirens* foliage. Changes in composition during maturation. *Journal of Agricultural and Food Chemistry*, 29(2), 324–326.

[2] Hall, G. D., & Langenheim, J. H. (1986). Within-tree spatial variation in the leaf monoterpenes of *Sequoia sempervirens*. *Biochemical Systematics and Ecology*, 14(6), 625–632.

[3] Hall, G. D., & Langenheim, J. H. (1987). Geographic variation in leaf monoterpenes of *Sequoia sempervirens*. *Biochemical Systematics and Ecology*, 15(1), 31–43.

[4] Espinosa-García, F. J., & Langenheim, J. H. (1991). Effects of sabinene and γ-terpinene from coastal redwood leaves acting singly or in mixtures on the growth of some of their fungus endophytes. *Biochemical Systematics and Ecology*, 19(8), 643–650.

[5] Taha, K. F., & Shakour, Z. T. A. (2016). Chemical composition and antibacterial activity of volatile oil of *Sequoia sempervirens* (Lamb.) grown in Egypt. *Medicinal & Aromatic Plants*, 5(3), 1–5.

[6] Marzouk, M., Farid, M. M., Fathy, H. M., Abd El-Shakour, Z. T., & Sokkar, N. M. (2023). GC/MS analysis and secondary metabolites determination with antioxidant and anticancer evaluation of *Sequoia sempervirens* (D. Don) Endl. through micropropagated cultures. *Egyptian Journal of Chemistry*, 66(11), 211–223.

[7] Lamarti, A., Badoc, A., Deffieux, G., & Carde, J.-P. (1994). Biogenèse des monoterpènes. III - Monoterpènes synthétases. *Bulletin de la Société de Pharmacie de Bordeaux*, 133, 100–118.