



Article

Phenolic Compounds as Unambiguous Chemical Markers for the Identification of Keystone Plant Species in the Bale Mountains, Ethiopia

Bruk Lemma ^{1,2,*} , Claudius Grehl ^{1,3}, Michael Zech ^{1,4}, Betelhem Mekonnen ^{1,5}, Wolfgang Zech ⁶, Sileshi Nemomissa ⁷, Tamrat Bekele ⁷ and Bruno Glaser ¹ 

¹ Institute of Agronomy and Nutritional Sciences, Soil Biogeochemistry, Martin Luther University Halle–Wittenberg, Von–Seckendorff–Platz 3, D–06120 Halle, Germany

² Ethiopian Biodiversity Institute, Forest and Rangeland Biodiversity Directorate, P.O. Box 30726 Addis Ababa, Ethiopia

³ Institute of Computer Science, Bioinformatics, Martin Luther University Halle–Wittenberg, Von Seckendorff-Platz 1, 06120 Halle (Saale), Germany

⁴ Institute of Geography, Technical University of Dresden, Helmholtzstrasse 10, D–01062 Dresden, Germany

⁵ Department of Urban Agriculture, Misrak Polytechnic College, P.O. Box 785, Addis Ababa, Ethiopia

⁶ Institute of Soil Science and Soil Geography, University of Bayreuth, Universitätsstrasse 30, D–95440 Bayreuth, Germany

⁷ Department of Plant Biology and Biodiversity Management, Addis Ababa University, P.O. Box 3434 Addis Ababa, Ethiopia

* Correspondence: bruklemma@gmail.com

Received: 17 June 2019; Accepted: 12 July 2019; Published: 16 July 2019



Abstract: Despite the fact that the vegetation pattern and history of the Bale Mountains in Ethiopia were reconstructed using pollen, little is known about the former extent of *Erica* species. The main objective of the present study is to identify unambiguous chemical proxies from plant-derived phenolic compounds to characterize *Erica* and other keystone species. Mild alkaline CuO oxidation has been used to extract sixteen phenolic compounds. After removal of undesired impurities, individual phenols were separated by gas chromatography and were detected by mass spectrometry. While conventional phenol ratios such as syringyl vs. vanillyl and cinnamyl vs. vanillyl and hierarchical cluster analysis of phenols failed for unambiguous *Erica* identification, the relative abundance of coumaryl phenols (>0.20) and benzoic acids (0.05–0.12) can be used as a proxy to distinguish *Erica* from other plant species. Moreover, a Random Forest decision tree based on syringyl phenols, benzoic acids (>0.06), coumaryl phenols (<0.21), hydroxybenzoic acids, and vanillyl phenols (>0.3) could be established for unambiguous *Erica* identification. In conclusion, serious caution should be given before interpreting this calibration study in paleovegetation reconstruction in respect of degradation and underground inputs of soil organic matter.

Keywords: paleoclimate; pollen; paleovegetation; oxidation; phenols; *Erica*; biomarkers; machine learning

1. Introduction

The present vegetation cover in mountain ecosystems is the result of subsequent historical and ecological processes primarily influenced by climate and human [1,2]. However, natural versus human impact on ecosystem dynamics at higher elevation is difficult to quantify and is still speculative [1,3]. To reconstruct and draw full pictures of environmental changes over the past centuries, Ethiopia is an ideal place for several reasons. Firstly, it is the origin of *Homo sapiens* and many other fauna and

flora. Secondly, it comprises diverse geographical zones attributed to various climatic features [4,5]. The Bale Mountains are one of the Eastern Afromontane Biodiversity Hotspots, which encompass many endemic and endangered species of Ethiopia [6,7]. Diverse topographical features of the Bale Mountains are marked by altitudinal vegetation zones. The ecosystem of the Bale Mountains is divided into three main declivities and five distinct and unique habitats. The Southern declivity is demarcated by moist Afromontane Haremma forest and a prominent Ericaceous belt whereas, the central plateau is dominated by the Afro-alpine vegetation, and the Northern slopes encompass dry Afromontane forest, as well as Gaysay grassland [6,8,9]. The afro-alpine regions of the Sanetti Plateau (3800–4377 m a.s.l.) are characterized by tussock grasses (mainly *Festuca* spp.), dwarf shrubs, and herbaceous plants (e.g. *Alchemilla haumanii*, *Helichrysum splendidum* and *Lobelia rhynchopetalum*) [8–10]. The prominent Ericaceous belt dominant above 3300 m a.s.l. shows different stages of post-fire succession and remains continuous up to 3800 m a.s.l. However, it becomes patchy on the Sanetti Plateau above 3800 m a.s.l. [9]. The characteristic species of the Northern slopes are *Hagenia abyssinica*, *Hypericum revolutum*, *Juniperus procera*, and *Podocarpus falcatus* [6,8].

In the Bale Mountains, the contemporary distribution of the *Erica* vegetation in specific and of the afro-alpine biodiversity in general is the legacy of climate change [2] or human-induced driving factors like fire and grazing [9] or a combination of both. Hitherto, this globally important ecosystem is still increasingly under threat of climate change and human impacts [11]. The vegetation history of the Bale Mountains was investigated and reconstructed using pollen records from lacustrine sediments and peat deposits [2,12–14]. The findings showed that *Erica* shrubs inhabited different altitudinal zones during different time periods. During the early Holocene, *Erica* shrubs had extended up to the Sanetti plateau in response to increased moisture and temperature [2]. *Erica* shrubs and forest had retreated to lower altitude in response to lower rainfall at about 6000 cal years BP and onwards [2]. Presently, *Erica* forming an Ericaceous belt is located between 3300 to 3800 m. a.s.l. [9].

The biggest deterrent of pollen-based studies depends on pollen preservation and can be biased by the dispersal potential of pollinators [3,15,16]. To overcome such drawbacks, we tested, for the first time, the chemotaxonomic roles of phenol biomarkers in the dominant plant species of the Bale Mountains. Chemical biomarkers are not influenced by a dispersal of parent material and can be extracted from sedimentary archives and soils and reflect more of the standing vegetation [17]. Therefore, they offer the potential to complement pollen-based vegetation reconstructions and to detect vegetation at a higher temporal and spatial resolution.

Polyphenols are the most widespread class of secondary metabolites, and ubiquitous in their distribution [18]. Phenolic compounds are a major component of terrestrial vascular plants [19] and may have lignin and non-lignin-derived biochemical structures. Secondary metabolites, mainly phenolic compounds, are important for cell wall structural formation, defense against pathogen, and environmental stress as well as essential for survival of vascular plants [20,21]. Plant phenolic compounds also exhibited significant qualitative and quantitative variation at different genetic level between and within species and clones [22–24]. They are extensively used in botanical chemosystematic studies [20]. The chemotaxonomic values of polyphenolic biomolecules such as tannins has been well recognized in the plant kingdom and the distribution of hydrolysable tannins used as chemotaxonomic markers in the family *Rosaceae* [25,26] and chestnut (family: *Fagaceae*) [27]. Stilbenoids, family of polyphenols, due to its limited but heterogeneous distribution in the plant kingdom also played a significant taxonomic role in the plant families such as *Vitaceae*, *Gnetaceae*, and *Fabaceae* [28]. Notably gossypetin (flavonol) are useful taxonomic marker for distinguishing Ericaceous species/genera (family: *Ericaceae*) [29]. The absence of leucoanthocyanidins, ellagic acid and phenyl-trihydroxylated characterized the genus *Pittosporum* (Pittosporaceae) [30]. Even if there is a need to encourage phenols application in chemotaxonomic studies/plant phylogeny, as more studies have been completed, some inconsistency has been revealed [31]. For instance the polyphenolic compound differences among the various tribes of *Anacardiaceae* are not very pronounced [32].

Nevertheless, little has been done to chemotaxonomically characterize *Erica* and other keystone species specifically in the Bale Mountains and generally in other similar geographical regions using phenols as a tool for paleovegetation reconstruction. Apart from the chemotaxonomic significance of phenols, the antioxidant and antibacterial activities of phenolic compounds by Guendouze-Bouchefa et al. [33], as well as the role of phenolic compounds as indicators of presence of metals in the leaves of different genera in the family *Ericaceae* by Marquez-Garcia et al. [34,35] were described previously. The antioxidant strength of *Erica* species depend upon the total abundance of phenolic compounds [34]. The abundance of phenolic compounds in *Festuca vallesiaca* were also used as indicators of fodder quality. High contents of total phenolic in the grass species are unfavorable for digestion [36]. Polyphenols obtained from flowers of *Helichrysum* (family: *Asteraceae*) species are used by some folks as ethno-medicine [37–40]. The affinity bounds in plant systematic classification using profiles of phenolic compounds were computed and compared, commonly using hierarchal cluster analysis (HCA) and principal component analysis (PCA) [20,41–43].

While the overall aim of our research is to understand how humans benefited from and re-shaped the African high altitude ecosystem during Quaternary climate changes, this study mainly focuses on identifying the potential of plant derived phenolic compounds as a proxy for unambiguous identification of *Erica* and other keystone species in the Bale Mountains. More specifically, this research attempts to addresses the following questions: (i) Do phenol biomarkers allow a chemotaxonomic differentiation of the contemporary dominant plant species of the Bale Mountains? (ii) Which implications have to be drawn from those results for planned paleovegetation reconstructions in the study area, e.g., concerning the reconstruction of the former extent of *Erica* stands?

2. Results and Discussion

2.1. Distribution and Diversity of Phenols

A typical gas chromatogram of phenolic compounds released after mild alkaline CuO oxidation illustrates the compositional complexity of a typical plant sample (*Lobelia rhynchopetalum*; Figure 1). The mean weighted sum of phenols content of *Alchemilla haumanii*, *Erica* spp., *Helichrysum splendidum*, *Kniphofia foliosa*, *Lobelia rhynchopetalum*, and *Festuca abyssinica* were 18, 16, 22, 6, 22, and 51 g kg⁻¹ TOC, respectively (Table S1).

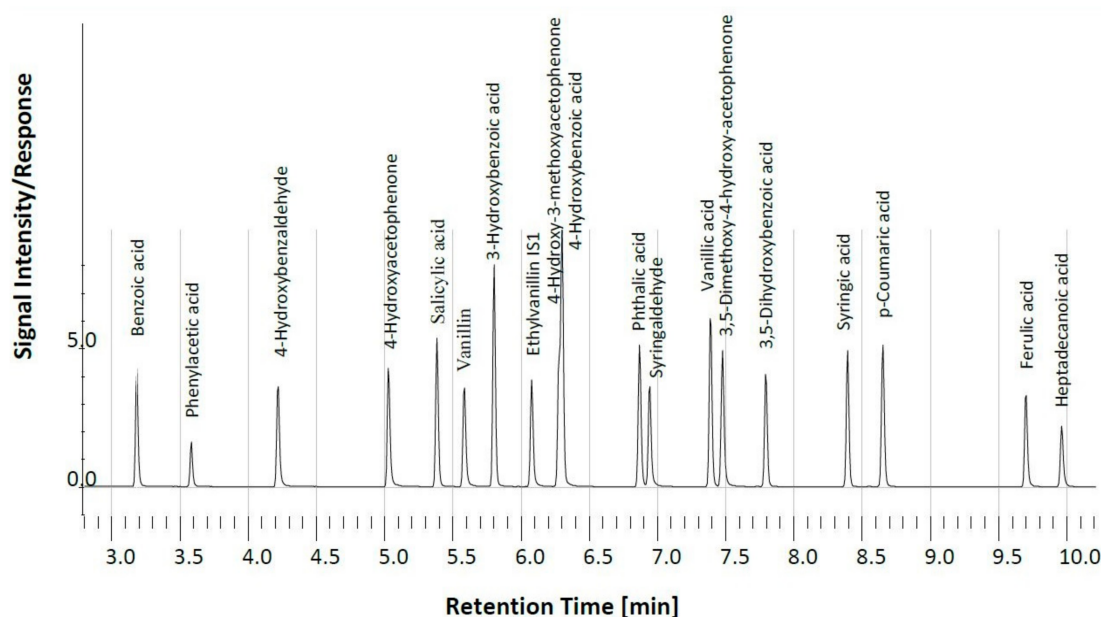


Figure 1. Typical gas chromatogram of phenolic compounds released after alkaline CuO oxidation of *Lobelia rhynchopetalum*.

Festuca abyssinica (Poaceae) is characterized by high abundance of coumaryl phenols (*p*-coumaric acid, ferulic acid) and syringyl phenols (syringic acid, 3,5-Dimethoxy-4-hydroxyacetophenone; Figure 2) and low abundance of vanillyl phenols and benzoic acids (Table 1). Similarly, the phenolic compounds from *Festuca vallesiaca* species investigated in Serbia accounts for low benzoic acid derivatives and high coumaryl phenols (*p*-coumaric acid, ferulic acid). The total phenolics extracted from *Festuca vallesiaca* using Soxlet extraction and later quantified by HPLC is 26.1 mg g⁻¹ [36]. In the grass taxon, most of phenol-related information was provided by coumaryl phenols (*p*-coumaric acid, ferulic acid) due to ample presence of the enzyme phenyl (thyrosine) ammonia-lyase and it is known for its high biological activity [20,36,44]. The abundance of coumaryl phenols in our *Poaceae* samples holds one third of the total phenols. Likewise, Hedges and Mann [45] stated that non woody vascular angiosperm plants have characteristic products of coumaryl phenols and it accounts for approximately one third of the total phenolic compounds. *Lobelia rhynchopetalum* and *Kniphofia foliosa* have high abundance of benzoic acid and 4-hydroxy-3-methoxyacetophenone (Table 1). The total phenolic contents of *Lobelia chinensis* measured by the Folin-Ciocalteu methods is 4.7 mg GAE g⁻¹ [46].

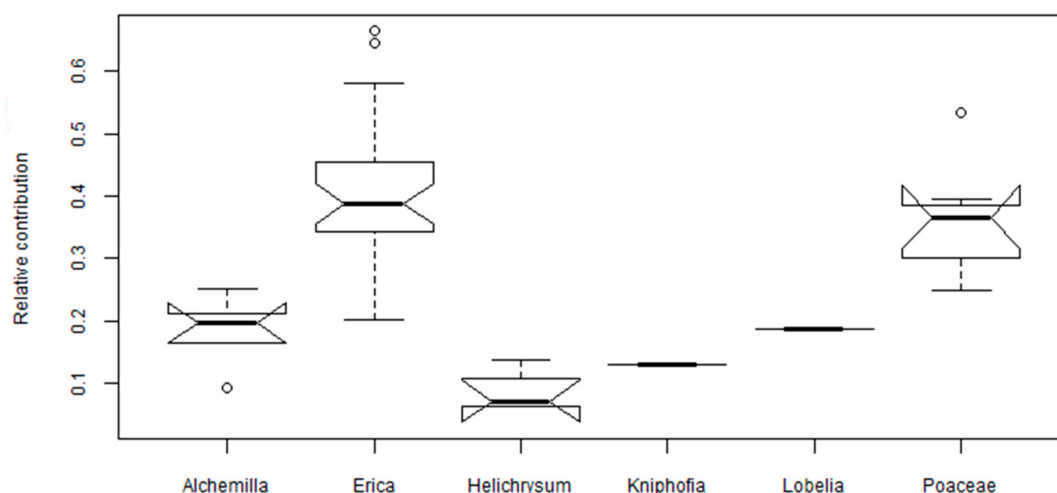
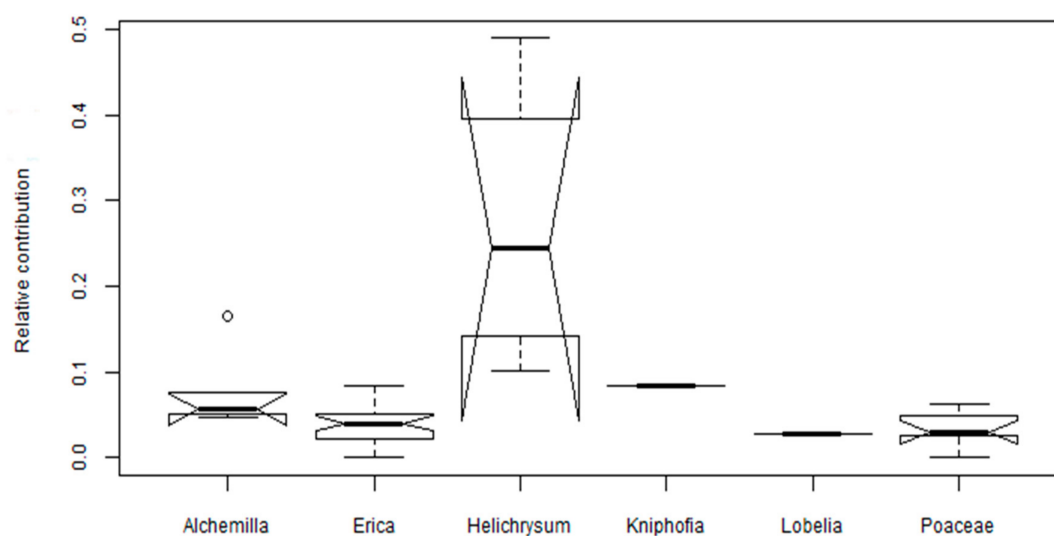


Figure 2. Notched box plots illustrating the abundance of coumaryl phenols in the dominant plant species of the Bale Mountains. The box plots indicate the median (solid line between the boxes), interquartile range (IQR) with upper (75%) and lower (25%) quartiles and possible outlier (white circles). The notches display the confidence interval around the median within $\pm 1.57 * IQR / \sqrt{n}$.

The sum of *p*-hydroxy phenols in the leaves unambiguously identifies *Helichrysum splendidum* from the other dominant plant species of the Bale Mountains (Figure 3). Qualitative analysis by thin layer chromatography (TLC) of the inflorescence of *Helichrysum* showed the presence of phenolic acids such as syringic, coumaric, and *p*-hydroxybenzoic [40]. The total phenolic contents extracted from different species of *Helichrysum* in Eastern Anatolia, Turkey ranges between 72 to 146 mg GAE g⁻¹ [39]. On the other hand, *Erica* species were characterized by high coumaryl (Figure 2) and low *p*-hydroxy phenols (Figure 3). The genus *Erica* (family: *Ericaceae*) is characterized by having polyphenols (*p*-coumaric acid derivate, vanillic acid, cinnamic acid derivate, and caffeic acid) and flavonoids [33]. The total phenolic compounds determined via Folin-Ciocalteu method and analysis by HPLC for *Erica arborea* and *Erica multiflora* are 71 and 68 mg GAE g⁻¹, respectively [33]. In the genus *Erica* among the flavonoids, dihydromyricetin 3-*O*- α -L-rhamnopyranoside is the most important chemotaxonomic marker [47]. Each of the dominant plant species on the Bale Mountains is profiled by a different abundance of phenolic compounds (Table 1).

Table 1. Summary of the relative abundance of individual phenols profiled in each dominant plant species. Key: ↑ (higher) and ↓ (lower).

Relative Contribution	<i>Alchemilla</i>	<i>Erica</i>	<i>Helichrysum</i>	<i>Kniphofia</i>	<i>Lobelia</i>	<i>Poaceae</i>
Vanillyl phenols					↑	↓
Vanillin						
Vanillic_acid						
_4_hydroxy_3_methoxyacetophenone						
Syringyl phenols	↑				↓	↑
Syringaldehyde						
Syringic_acid						
_3_5_dimethoxy_4_acetophenone						
Coumaryl phenols		↑	↓			↑
p_coumaric_acid						
Ferulic_acid						
p_hydroxy_phenols		↓	↑			
_4_hydroxybenzaldehyde						
_4_hydroxybenzacetophenone						
Benzoic_acids				↑	↑	↓
Benzoic_acid						
Salicylic_acid						
Phthalic_acid						
Hydroxy_benzoic_acids	↑		↓			↑
_3_hydroxybenzoic_acid						
_4_hydroxybenzoic_acid						
_3_5_Dihydroxy_benzoic_acid						

**Figure 3.** Notched box plots show the relative abundance of *p*-hydroxy phenols in the dominant plant species of the Bale Mountains. The box plots indicate the median (solid line between the boxes), interquartile range (IQR) with upper (75%) and lower (25%) quartiles and possible outlier (white circles). The notches display the confidence interval around the median within $\pm 1.57 * IQR / \sqrt{n}$.

Even though *Kniphofia foliosa* has lower phenolic content, the quantitative values of Shannon-Wiener phenol diversity index ($H = -\sum (p_i) [\ln(p_i)]$) indicated that *Kniphofia foliosa* has evenly distributed phenolic compounds. By contrast, *Helichrysum splendidum* has less evenly distributed phenolic compounds (Table 2). Innovative techniques and widespread occurrence of secondary metabolites like phenolic profiles in the plant diversity allow low taxonomic level chemosystematic studies [18,20].

Table 2. Phenol Diversity Index.

Species	Diversity Index (H)
<i>Alchemilla haumannii</i>	2.28
<i>Erica</i> spp.	2.19
<i>Helichrysum splendidum</i>	1.63
<i>Kniphofia foliosa</i>	2.47
<i>Lobelia rynchopetalum</i>	2.09
<i>Festuca abyssinica</i>	2.17

2.2. Cluster Analysis of Phenolic Compounds

The cluster analysis computed for 47 plant samples grouped the dominant plant species into six groups, as shown in Figure 4. In the cluster, *Poaceae* and *Lobelia* species are grouped independently (there are two independent *Poaceae* groups), while the other dominant species *Erica*, *Kniphofia*, *Helichrysum*, and *Alchemilla* were clustered together into two different groups (Figure 4). Therefore, no unambiguous identification of *Erica* is possible, but identify the former extent of the distribution of *Poaceae* in the Bale Mountains using alkaline CuO phenol products might be feasible (Figure 4). Previously, some studies used phenolic compounds to systematically characterize grasses even at lower taxonomic level [18]. However, phenols are environmentally labile secondary metabolites. As a result, caution should be given before extraction and standardizing the growing condition and growth stage of the sample materials [20,21,48]. Thus far, our cluster analysis did not justify which phenol type is the most determining characteristic for the identification of the dominant plant species, one from the other. The only obvious independent group is the *Lobelia* one, and *Poaceae* is also somewhat distinct. Therefore, the ambiguity is less clear for all the rest.

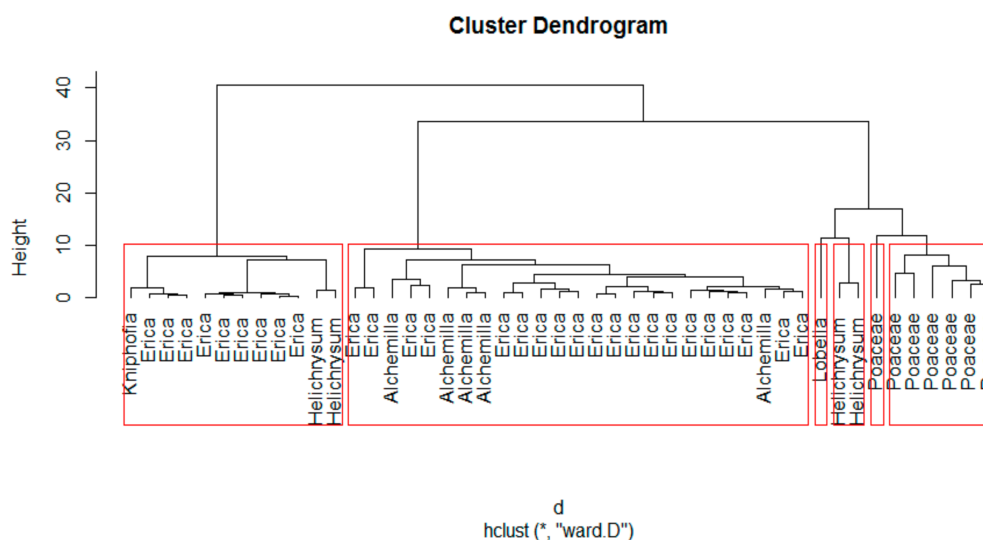
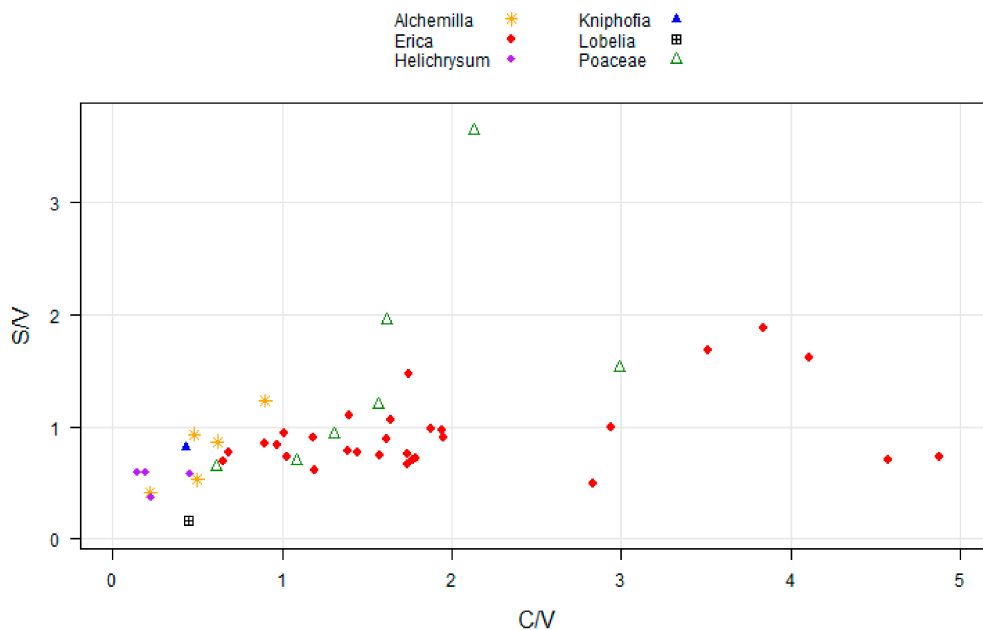


Figure 4. Agglomerated Hierarchical Cluster analysis of the dominant plant species of the Bale Mountains based on the abundance of phenolic compounds (g kg^{-1} TOC). Euclidean measurement of distances and Ward method for linkage calculation applied to cluster a normalized dataset (Z-score normalization, $Z = (X - \text{mean}(X)) / \text{sd}(X)$).

2.3. Biomarker Identification

The ratios of syringyl over vanillyl (S/V) and cinnamyl over vanillyl (C/V) phenols, which were used previously as a source proxy did not turn out to be useful in the present study to categorize the dominant plant species (Figure 5). The ratios of S/V and C/V allowed us to distinguish organic matter sources in soil and sediments [45]. However, these ratios cannot provide profound results of plants belonging to the same taxon [49]. The limitation of S/V and C/V ratios to pinpoint the source of

organic matter in soils and sediments of the same taxon were described in detail by Thevenot [49]. Here, we also proved that the ratios of S/V and C/V were unable to characterize modern plants of the same taxon at least in our study area (Figure 5). The drawback of those ratios could be associated with the overlap between plant molecular signatures [49,50]. The sampled woody vegetation (e.g. *Erica*, *Helichrysum*) in the present study exhibited C/V ratios higher than zero. However, Hedges and Mann [45] stated that woody gymnosperm and angiosperm plants are characterized by C/V ratio nearly equal to zero.



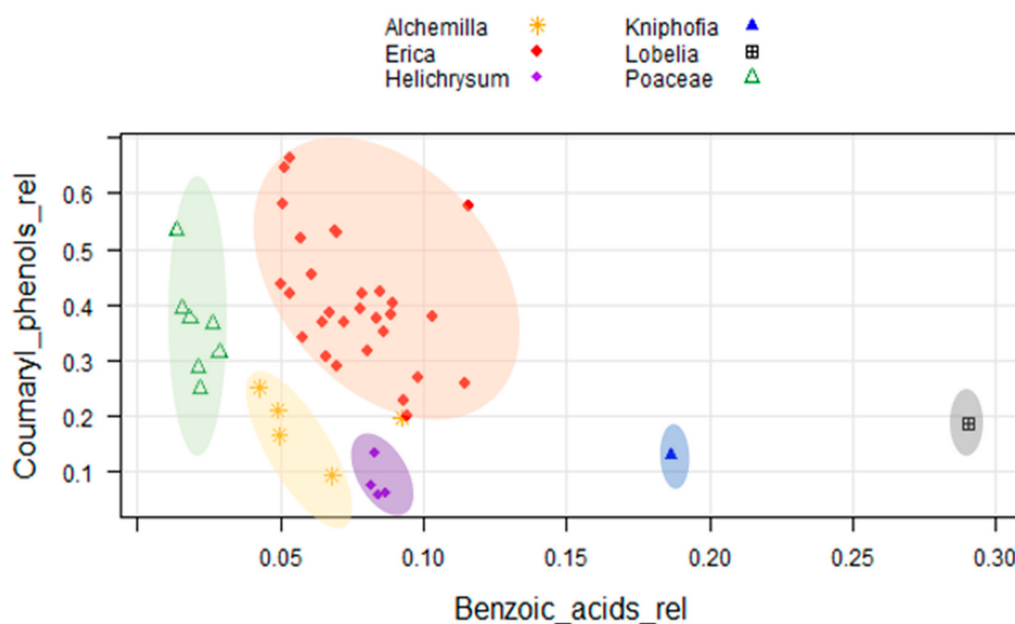


Figure 6. Two-dimensional plot of the relative contribution of the sum of coumaryl phenols (*p*-coumaric acid + ferulic acid) and benzoic acids (benzoic acid + salicylic acid + phthalic acid) in the dominant plant species under study.

Three different machine-learning algorithms namely Support Vector Machine, Random Forest, and Recursive Partitioning were tested to unambiguously identify *Erica* vegetation based on our phenolic compounds dataset. Among the tested machine learning approaches, Random Forest performed best. The proposed algorithm has been shown to be better in terms of various performance indicators like accuracy (Figure S1A) and F1-score (Figure S1B). In the experiment using relative phenols dataset, the key variable for chemotaxonomic classification of the contemporary species in the Bale Mountains are given in Figure 7. Among them, benzoic acids (<0.06) and coumaryl phenols (0.21) were the most decisive variables to identify *Erica* from the other species (Figures 6 and 8). The cross-validation of the used model is shown in Figure S3.

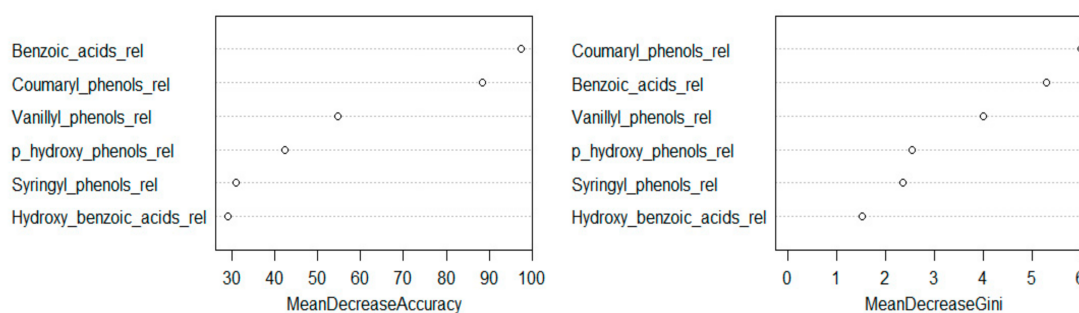


Figure 7. Set of relative phenols variables used for chemotaxy ordered by their importance as estimated by the Random Forest model. 32 training sample sets and 10,000 number of trees were used to compute the model.

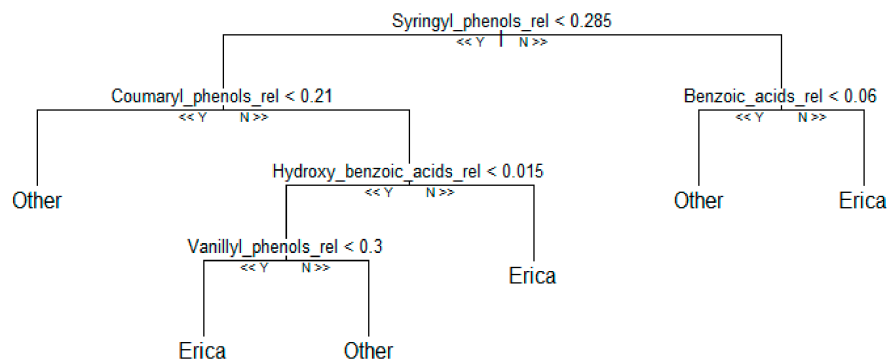


Figure 8. Representative decision tree of Random Forest model using relative phenols contribution. 32 training sample sets and 10,000 number of trees were used to compute the model.

3. Material and Methods

3.1. Study Area

3.1.1. Geomorphological Setting

The Bale Mountains belong to the Bale-Arsi mountain massif located 400 km Southeast of Addis Ababa, Ethiopia [6]. They are situated between 39°03' to 40°00' longitude (E) and 6°29' to 7°10' latitude (N) (Figure 9). Lava mainly outpouring 40 million years ago resulted in the formation of mountains and ridges up to 4377 m a.s.l. (Mohr, 1963). The most extensive high altitude plateau (Sanetti Plateau) in Africa above 3000 m. a.s.l., comprised within the Bale Mountains National Park, extends to about 2,600 km² [2,5,7,11]. The Bale Mountains are topographically divided into three major declivities: The Northern slopes (3000–3800 m. a.s.l.), the central plateau (3800–4377 m a.s.l), and the Southern Haremma escarpment (1400–3800 m. a.s.l.). The Sanetti Plateau is located in the center, on which the second highest peak of the country Mt. Tulu Dimtu (4377 m a.s.l.) rests. The northern slope of the mountains is dissected by the Togona Valley, which descends gently towards the extensive Arsi Plateau and further down to the Great Rift Valley lowlands, which divide the country into two parts. The Southern slope includes the steep Haremma escarpment and goes down to the surrounding lowland at about 1400 m a.s.l. [5,7]. Repeated glaciations of the high altitudes created typical features of glaciated landscapes with moraines and glacial lake [2,51]. Basalt and rhyolite are typical parent rocks [52] of the dominant Cambisols and Leptosol at the Sanetti Plateau [9,10,53].

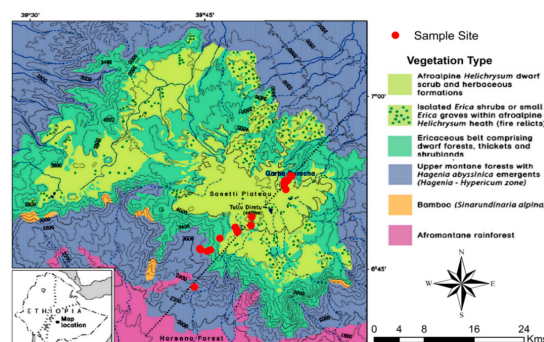


Figure 9. Map of the Bale Mountains showing the vegetation zones and study sites along the Northeast and Southwestern transect (modified after Miehe and Miehe, 1994). Leaves of the dominant vegetation within the Ericaceous and afro-alpine belts were sampled during the dry season in February 2015 and 2017.

3.1.2. Climate and Biota

The climatic conditions of the Bale Mountains depend on orography and are vulnerable to extreme climatic conditions over the past years [2]. The mean annual minimum temperature in the mountainous region ranges between 0.6 to 10 °C with frequent frost during winter season and mean annual maximum temperature ranges between 6–12 °C. The mean annual temperature in Dinsho headquarter (3170 m a.s.l.) is 11.8 °C [5,54]. Precipitation in the Bale Mountains is influenced by the shift of the Inter Tropical Convergence Zone (ITCZ) resulting in long rainy (March–September) and short dry (October–February) seasons [55]. The Southern slopes of the Bale Mountains receive high annual rainfall (1000–1500 mm year⁻¹) as compared to the Northern declivity (800–1000 mm year⁻¹) [5,53,54]. Moisture reaching the Bale Mountains originates from the Red Sea, the Mediterranean Sea, the Indian Ocean, and the Atlantic Ocean [9,56,57].

The Bale Mountains National Park and its surrounding are home for more than 1300 vascular flowering plant species and 50 mammal species [7,58]. The composition of enormous endemic and endangered fauna and flora labeled the Bale Mountains as one of the biodiversity hotspot areas of the world. The distribution and abundance of the contemporary vegetation in the Bale Mountains are shaped partly by human intervention and often emphasizes the landforms and paleoclimate [51]. Diverse topographical variability in the Bale Mountains shows altitudinal zonation of vegetation. In broad terms, floristically the mountains are divided into three topographical regions. The southern declivity is delineated by moist Afromontane forest, Ericaceous belt and Afro-alpine vegetation, whereas the central plateau is dominated by the Afro-alpine vegetation and the northern slope encompasses dry Afromontane forest, Gaysay grassland, Ericaceous belt, and afro-alpine zones [6,8,9]. The afro-alpine regions of the Sanetti Plateau (3800–4377 m a.s.l.) are characterized by tussock grassland and dwarf shrubs and herbaceous (e.g., *Alchemilla haumannii*, *Helichrysum splendidum*, and *Lobelia rhynchopetalum*) plants. The Ericaceous belt (3300–3800 m a.s.l.) comprises forest, thickets, and scrublands of *Erica trimera* and *Erica arborea* with mosses and grasses dominating in the ground layer. The Harena forest (1500–3300) is a natural remnant of a moist Afromontane forest dominated by broadleaved evergreen trees and clustered floristically into two different classes. Around 1500–2300 m a.s.l., the forest is dominated by *Podocarpus falcatus* associated with *Croton macrostachyus*, *Pouteria adolfi-friederici*, *Syzygium guineense*, *Warburgia ugandensis*, and worth noting *Coffea arabica*. Between 2300–3200 m a.s.l. *Arundinaria alpina*, *Hagenia abyssinica*, *Hypericum revolutum*, *Erythrina brucei*, *Prunus africana*, and *Schefflera volkensii* are the dominant plants [2,7–9,53]. Furthermore, the Bale Mountains are center for faunal diversity and holds the highest rate of animal endemism for a terrestrial habitat anywhere in the world [7]. The mountains are known by its flagship mammals like Mountain Nyala (*Tragelaphus buxtoni*), Ethiopian Wolf (*Canis simensis*) and Giant mole rat (*Tachyoryctes macrocephalus*). Also about 180 bird species and 14 amphibian species are inhabiting the Bale Mountains [6,7].

3.2. Sample Collection

25 and 22 (total of 47) samples (leaves and twig) of locally dominant plant species were collected in February 2015 and 2017, respectively along a SW and a NE transect (ranging from 2550 to 4377 m a.s.l. and 3870 to 4134 m a.s.l., respectively) (Figure 9). Samples comprise *Erica trimera* and *Erica arborea* ($n = 29$), *Alchemilla haumannii* Rothm. ($n = 5$), *Festuca abyssinica* Hochst. ex A. Rich. ($n = 7$), *Helichrysum splendidum* (Thunb.) Less. ($n = 4$), *Kniphofia foliosa* Hochst. ($n = 1$) and *Lobelia rhynchopetalum* Hemsl. ($n = 1$). All samples were air-dried in the Soil Store Laboratory of the National Herbarium, Department of Plant Biology and Biodiversity Management, Addis Ababa University. Samples were finely grinded, homogenized, and subjected to further biogeochemical analysis at the laboratories of the Soil Biogeochemistry, Martin Luther University of Halle–Wittenberg.

3.3. Analysis of Phenolic Compounds Released after Alkaline CuO Oxidation

Phenolic compounds were extracted from 35 mg of dried plant leave and twig samples using the mild alkaline cupric oxide (CuO) oxidation method developed by Hedges and Ertel [59] as later modified by Goñi and Hedges [60]. Briefly, the samples were transferred into Teflon digestion tubes together with 100 mg of $(\text{NH}_4)_2\text{Fe}(\text{SO}_4)_2 \cdot 6\text{H}_2\text{O}$, 500 mg of CuO, 50 mg of $\text{C}_6\text{H}_{12}\text{O}_6$, 1 mL of ethylvanillin solution (100 mg L^{-1}) as internal standard 1 and 15 mL of 2M NaOH and digested at $170 \text{ }^\circ\text{C}$ for two hours under elevated pressure. Reaction products were cooled overnight and transferred into centrifuge tubes. Then the phenolic compounds were purified by adsorption on C18 columns and desorbed by ethylacetate and concentrated under a stream of nitrogen gas for 30 min. Residue was dissolved in 1 mL phenylacetic acid (PAA), as internal standard 2 to determine the recovery of ethylvanillin before derivatization [61,62]. Finally, the samples were derivatized using 200 μL of BSTFA and 100 μL of pyridine. Oxidation products of phenolic compounds were quantified using a Gas Chromatograph coupled to a Mass Spectrometer (SHIMADZU, GC-MS-QP2010, Kyoto, Japan).

3.4. Data Analysis

After recovery correction, the content of each of sixteen phenolic compounds (in g kg^{-1} TOC) was calculated and groups of phenolic derivative of mild alkaline CuO oxidation products were calculated according to Equations (1)–(6).

$$p\text{-hydroxy} = 4\text{-Hydroxybenzaldehyde} + 4\text{-Hydroxyacetophenone} + 4\text{-Hydroxybenzoic acid} \quad (1)$$

$$\text{Vanillyl (V)} = \text{vanillin} + 4\text{-hydroxy-3-methoxyacetophenone} + \text{vanillic acid} \quad (2)$$

$$\text{Syringyl (S)} = \text{syringaldehyde} + 3,5\text{-dimethoxy-4-hydroxyacetophenone} + \text{syringic acid} \quad (3)$$

$$\text{Coumaryl (C)} = p\text{-coumaric acid} + \text{ferulic acid} \quad (4)$$

$$\text{Benzoic acids} = \text{benzoic acid} + \text{salicylic acid} + \text{phthalic acid} \quad (5)$$

$$\text{Hydroxy-Benzoic-Acids} = 3\text{-hydroxybenzoic acid} + 4\text{-hydroxybenzoic acid} + 3,5\text{-dihydroxybenzoic acid} \quad (6)$$

3.4.1. Hierarchical Clustering and PCA

Calibrated and normalized data sets of the analytical results were subjected to Principal Component Analysis (PCA, Figure S2) and hierarchical clustering (AHC) using Euclidean measurement of distances and Ward method for linkage calculation to examine taxonomic differences based on phenolic compounds. Z-score normalization method, $Z = (X - \text{mean}(X)) / \text{sd}(X)$, were used to standardize data. Furthermore, descriptive statistics (notched box plot and scatter plot) were performed to identify an unambiguous biomarker using R free software version 3.6.0 [63].

3.4.2. Machine Learning

We examined three different machine learning algorithms (Random forest, Support Vector Machines, and Recursive Partitioning) regarding their accuracy (Figure S1A) and F1 score (Figure S1B) based on different proportions of training and prediction dataset with five replications. Due to the reliable performance of Random forest classification of *Erica* vs. non-*Erica* species, we looked for the most important features in a 2/3 to 1/3 training against prediction setting. To verify the reliability of the used model, cross-validation has been performed (Figure S3).

4. Conclusions

Unambiguous identification of chemical proxies using mild alkaline CuO oxidation products of phenols were not simple to apply on plants of the same taxon growing along the SW and NE exposed transects in the Bale Mts. Nonetheless, leaves and twigs of the woody *Erica* shrubs can be distinguished from leaves of the present day dominant herbaceous Poaceae and other lower plants by the relative abundance of coumaryl phenols and benzoic acids. However, as it is known that coumaryl phenols are preferentially degraded in soils, these proxies are not suitable for the evaluation of sites covered in former times by *Erica* (reconstruction of palaeovegetation in soils and sediments). To avoid such ambiguities in future studies, identification of other suitable proxies such as tannin-derived phenols and terpenoids is recommended.

Supplementary Materials: The following are available online at <http://www.mdpi.com/2223-7747/8/7/228/s1>, Figure S1: Comparison of Accuracy (A) and F1 Score (B) of Support Vector Machine (SVM, blue), Random Forest (RF, red) and Recursive Partitioning (RP, grey) algorithm based on the level of relative phenols, in relation to the ratio of tested/total number of samples ($n = 47$) that has been split into test and training datasets, each model has been computed 5 times, the bars indicate the standard derivation between the prediction results of the models; Figure S2: Principal Component Analysis, PCA based on the relative phenol abundance in 47 leaf and twig samples from the Bale Mountains, shown are the first two principal components (PC1 and PC2). Dark red arrows indicate the direction of each vector of feature; Figure S3: Cross-validation of the model used; Table S1: Sum of weighted mean of phenolic compounds of each dominant plant species.

Author Contributions: Conceptualization: B.G., W.Z., M.Z.; sample collection: B.M., W.Z., B.L.; Biochemical analysis: B.L.; Data analysis, B.G., C.G., M.Z., B.L.; writing—original draft preparation, B.L., B.G.; writing—review and editing, S.N., T.B., C.G., M.Z., B.M., W.Z.

Funding: This research work was funded by the German Research Foundation within the DFG Research Unit ‘The Mountain Exile Hypothesis’ (DFG GL327/18–1, ZE844/10–1). Bruk Lemma expresses his gratitude to the Catholic Academic Exchange Services (KAAD, Germany) for a PhD fellowship. We acknowledge the financial support within the funding programme Open Access Publishing by the German Research Foundation (DFG).

Acknowledgments: The authors are grateful for the scientific and technical support obtained from Addis Ababa University, Ethiopian Biodiversity Institute and Ethiopian Wildlife Conservation Authority. Bruk Lemma is extended his gratitude to Heike Maennicke for kind assistance during laboratory work. We appreciate the comments from the anonymous reviewers to improve this manuscript.

Conflicts of Interest: The authors declare no conflict of interest. And, the funder had no role in the design of the study; in the collection, analyses, or interpretation of data; in the writing of the manuscript, or in the decision to publish the results.

References

1. Heiri, C.; Bugmann, H.; Tinner, W.; Heiri, O.; Lischke, H. A model-based reconstruction of Holocene treeline dynamics in the Central Swiss Alps. *J. Ecol.* **2006**, *94*, 206–216. [[CrossRef](#)]
2. Umer, M.; Lamb, H.F.; Bonnefille, R.; Lézine, A.M.; Tiercelin, J.J.; Gibert, E.; Cazet, J.P.; Watrin, J. Late Pleistocene and Holocene vegetation history of the Bale Mountains, Ethiopia. *Quat. Sci. Rev.* **2007**, *26*, 2229–2246. [[CrossRef](#)]
3. Jansen, B.; van Loon, E.E.; Hooghiemstra, H.; Verstraten, J.M. Improved reconstruction of palaeo-environments through unravelling of preserved vegetation biomarker patterns. *Palaeogeogr. Palaeoclimatol. Palaeoecol.* **2010**, *285*, 119–130. [[CrossRef](#)]
4. Terwilliger, V.J.; Eshetu, Z.; Colman, A.; Bekele, T.; Gezahgne, A.; Fogel, M.L. Reconstructing palaeoenvironment from $\delta^{13}\text{C}$ and $\delta^{15}\text{N}$ values of soil organic matter: A calibration from arid and wetter elevation transects in Ethiopia. *Geoderma* **2008**, *147*, 197–210. [[CrossRef](#)]
5. Tiercelin, J.J.; Gibert, E.; Umer, M.; Bonnefille, R.; Disnar, J.R.; Lézine, A.M.; Hureau-Mazaudier, D.; Travi, Y.; Keravis, D.; Lamb, H.F. High-resolution sedimentary record of the last deglaciation from a high-altitude lake in Ethiopia. *Quat. Sci. Rev.* **2008**, *27*, 449–467. [[CrossRef](#)]
6. Hillman, J.C. Conservation in Bale Mountains National Park, Ethiopia. *Oryx* **1986**, *20*, 89–94. [[CrossRef](#)]
7. Hillman, J.C. The Bale Mountains National Park area, Southeast Ethiopia, and its management. *Mt. Res. Dev.* **1988**, *8*, 253–258. [[CrossRef](#)]

8. Friis, I. Zonation of Forest Vegetation on the South Slope of Bale Mountains, South Ethiopia. *SINET Ethiop. J. Sci.* **1986**, *9*, 29–44.
9. Miehe, G.; Miehe, S. *Ericaceous Forests and Heathlands in the Bale Mountains of South Ethiopia: Ecology and Man's Impact*; T. Warnke Verlag: Hamburg, Germany, 1994; p. 206.
10. Hedberg, O. *Features of Afroalpine Plant Ecology*; Swedish Science Press: Uppsala, Sweden, 1964; p. 148.
11. Kidane, Y.; Stahlmann, R.; Beierkuhnlein, C. Vegetation dynamics, and land use and land cover change in the Bale Mountains, Ethiopia. *Environ. Monit. Assess.* **2012**, *184*, 7473–7489. [[CrossRef](#)]
12. Bonnefille, R.; Hamilton, A. Quaternary and Late Tertiary history of Ethiopian vegetation. *Symb. Bot. Ups.* **1986**, *26*, 48–63.
13. Bonnefille, R.; Mohammed, U. Pollen-inferred climatic fluctuations in Ethiopia during the last 3000 years. *Palaeogeogr. Palaeoclimatol. Palaeoecol.* **1994**, *109*, 331–343. [[CrossRef](#)]
14. Hamilton, A.C. *Environmental History of East Africa: A Study of the Quaternary*; Academic Press: London, UK, 1982; p. 328.
15. Hicks, S. When no pollen does not mean no trees. *Veg. Hist. Archaeobotany* **2006**, *15*, 253–261. [[CrossRef](#)]
16. Ortu, E.; Brewer, S.; Peyron, O. Pollen-inferred palaeoclimate reconstructions in mountain areas: Problems and perspectives. *J. Quat. Sci.* **2006**, *21*, 615–627. [[CrossRef](#)]
17. Glaser, B.; Zech, W. Reconstruction of climate and landscape changes in a high mountain lake catchment in the Gorkha Himal, Nepal during the Late Glacial and Holocene as deduced from radiocarbon and compound-specific stable isotope analysis of terrestrial, aquatic and microbial biomarkers. *Org. Geochem.* **2005**, *36*, 1086–1098. [[CrossRef](#)]
18. Ram Singh, C. Chemotaxonomy: A Tool for Plant Classification. *J. Med. Plants Stud. JMPS* **2016**, *90*, 90–93.
19. Ertel, J.R.; Hedges, J.I. The lignin component of humic substances: Distribution among soil and sedimentary humic, fulvic, and base-insoluble fractions. *Geochim. Cosmochim. Acta* **1984**, *48*, 2065–2074. [[CrossRef](#)]
20. Míka, V.; Kubáň, V.; Klejdus, B.; Odstrčilová, V.; Nerušil, P. Phenolic compounds as chemical markers of low taxonomic levels in the family Poaceae. *Plant Soil Environ.* **2005**, *51*, 506–512. [[CrossRef](#)]
21. Ramakrishna, A.; Ravishankar, G.A. Influence of abiotic stress signals on secondary metabolites in plants. *Plant Signal. Behav.* **2011**, *6*, 1720–1731.
22. Witzell, J.; Gref, R.; Na, T. Plant-part specific and temporal variation in phenolic compounds of boreal bilberry (*Vaccinium myrtillus*) plants. *Biochem. Syst. Ecol.* **2003**, *31*, 115–127. [[CrossRef](#)]
23. Nichols-Orians, C.M.; Fritz, R.S.; Clausen, T.P. The genetic basis for variation in the concentration of phenolic glycosides in *Salix sericea*: Clonal variation and sex-based differences. *Biochem. Syst. Ecol.* **1993**, *21*, 535–542. [[CrossRef](#)]
24. Hakulinen, J.; Julkunen-Tiitto, R.; Tahvanainen, J. Does nitrogen fertilization have an impact on the trade-off between willow growth and defensive secondary metabolism? *Trees* **1995**, *9*, 235–240. [[CrossRef](#)]
25. Okuda, T.; Yoshida, T.; Hatano, T.; Iwasaki, M.; Kubo, M.; Orime, T.; Yoshizaki, M.; Naruhashi, N. Hydrolysable tannins as chemotaxonomic markers in the Rosaceae. *Phytochemistry* **1992**, *31*, 3091–3096. [[CrossRef](#)]
26. Määttä-Riihinen, K.R.; Kamal-Eldin, A.; Törrönen, A.R. Identification and quantification of phenolic compounds in berries of *Fragaria* and *Rubus* species (family Rosaceae). *J. Agric. Food Chem.* **2004**, *52*, 6178–6187. [[CrossRef](#)]
27. Fernandez-Lorenzo, J.L.; Rigueiro, A.; Ballester, A. Polyphenols as potential markers to differentiate juvenile and mature chestnut shoot cultures. *Tree Physiol.* **1999**, *19*, 461–466. [[CrossRef](#)]
28. Rivière, C.; Pawlus, A.D.; Mérillon, J.-M. Natural stilbenoids: Distribution in the plant kingdom and chemotaxonomic interest in Vitaceae. *Nat. Prod. Rep.* **2012**, *29*, 1317–1333. [[CrossRef](#)]
29. Harborne, J.B.; Williams, C.A. A chemotaxonomic survey of flavonoids and simple phenols in leaves of the Ericaceae. *Bot. J. Linn. Soc.* **1973**, *66*, 37–54. [[CrossRef](#)]
30. Jay, M. Chemotaxonomic researches on vascular plants. XIX. Flavonoid distribution in the Pittosporaceae. *Bot. J. Linn. Soc.* **1969**, *62*, 423–429. [[CrossRef](#)]
31. Giannasi, D.E. Systematic aspects of flavonoid biosynthesis and evolution. *Bot. Rev.* **1978**, *44*, 399–429. [[CrossRef](#)]
32. Umadevi, I.; Daniel, M.; Sabnis, S.D. Chemotaxonomic studies on some members of Anacardiaceae. *Proc. Plant Sci.* **1988**, *98*, 205–208.

33. Guendouze-Bouchefa, N.; Madani, K.; Chibane, M.; Boulekbache-Makhlouf, L.; Hauchard, D.; Kiendrebeogo, M.; Stévigny, C.; Okusa, P.N.; Duez, P. Phenolic compounds, antioxidant and antibacterial activities of three Ericaceae from Algeria. *Ind. Crop. Prod.* **2015**, *70*, 459–466. [[CrossRef](#)]
34. Márquez-García, B.; Fernández, M.Á.; Córdoba, F. Phenolics composition in Erica sp. differentially exposed to metal pollution in the Iberian Southwestern Pyritic Belt. *Bioresour. Technol.* **2009**, *100*, 446–451. [[CrossRef](#)]
35. Márquez-García, B.; Fernández-Recamales, M.; Córdoba, F. Effects of cadmium on phenolic composition and antioxidant activities of Erica andevalensis. *J. Bot.* **2012**, *2012*, 936950. [[CrossRef](#)]
36. Djurdjević, L.; Mitrović, M.; Pavlović, P.; Perišić, S. Total phenolics and phenolic acids content in low (*Chrysopogon gryllus*) and mediocre quality (*Festuca vallesiaca*) forage grasses of Deliblato Sands meadow-pasture communities in Serbia. *Czech J. Anim. Sci.* **2005**, *50*, 54–59. [[CrossRef](#)]
37. Pljevljakušić, D.; Bigović, D.; Janković, T.; Jelačić, S.; Šavikin, K. Sandy everlasting (*Helichrysum arenarium* (L.) Moench): Botanical, chemical and biological properties. *Front. Plant Sci.* **2018**, *9*, 1–12. [[CrossRef](#)]
38. Babotă, M.; Mocan, A.; Vlase, L.; Crișan, O.; Ielciu, I.; Gheldiu, A.-M.; Vodnar, D.; Crișan, G.; Păltinean, R. Phytochemical analysis, antioxidant and antimicrobial activities of *Helichrysum arenarium* (L.) Moench and *Antennaria dioica* (L.) Gaertn. flowers. *Molecules* **2018**, *23*, 409. [[CrossRef](#)]
39. Albayrak, S.; Aksoy, A.; SaĖdiĖ, O.; Budak, Ü. Phenolic compounds and antioxidant and antimicrobial properties of *Helichrysum* species collected from eastern Anatolia, Turkey. *Turk. J. Biol.* **2010**, *34*, 463–473.
40. Sroka, Z.; Kuta, I.; Cisowski, W.; Dryś, A. Antiradical activity of hydrolyzed and non-hydrolyzed extracts from *Helichrysi inflorescentia* and its phenolic contents. *Z. Naturforschung C* **2004**, *59*, 363–367. [[CrossRef](#)]
41. Liu, Z.; Liu, Y.; Liu, C.; Song, Z.; Li, Q.; Zha, Q.; Lu, C.; Wang, C.; Ning, Z.; Zhang, Y.; et al. The chemotaxonomic classification of *Rhodiola* plants and its correlation with morphological characteristics and genetic taxonomy. *Chem. Cent. J.* **2013**, *7*, 118. [[CrossRef](#)]
42. Pan, Y.; Zhang, J.; Zhao, Y.-L.; Wang, Y.-Z.; Jin, H. Chemotaxonomic studies of nine Gentianaceae species from western China based on liquid chromatography tandem mass spectrometry and Fourier transform infrared spectroscopy. *Phytochem. Anal.* **2016**, *27*, 158–167. [[CrossRef](#)]
43. Cruz, A.V.M.; Ferreira, M.J.; Scotti, M.T.; Kaplan, M.A.; Emerenciano, V.P. Chemotaxonomic relationships in Celastraceae inferred from principal component analysis (PCA) and partial least squares (PLS). *Nat. Prod. Commun.* **2008**, *3*, 911–917. [[CrossRef](#)]
44. Chesson, A.; Stewart, C.S.; Wallace, R.J. Influence of Plant Phenolic Acids on Growth and Cellulolytic Activity of Rumen Bacteria. *Appl. Environ. Microbiol.* **1982**, *44*, 597–603.
45. Hedges, J.I.; Mann, D.C. The characterization of plant tissues by their lignin oxidation products. *Geochim. Cosmochim. Acta* **1979**, *43*, 1803–1807. [[CrossRef](#)]
46. Li, H.-B.; Wong, C.-C.; Cheng, K.-W.; Chen, F. Antioxidant properties in vitro and total phenolic contents in methanol extracts from medicinal plants. *LWT Food Sci. Technol.* **2008**, *41*, 385–390. [[CrossRef](#)]
47. Nazemiyeh, H.; Bahadori, F.; Delazar, A.; Ay, M.; Topçu, G.; Nahar, L.; Majinda, R.R.; Sarker, S.D. Antioxidant phenolic compounds from the leaves of *Erica arborea* (Ericaceae). *Nat. Prod. Res.* **2008**, *22*, 1385–1392. [[CrossRef](#)]
48. Stanković, M.; Ćurčić, S.; Zlatić, N.; Bojović, B. Ecological variability of the phenolic compounds of *Olea europaea* L. leaves from natural habitats and cultivated conditions. *Biotechnol. Biotechnol. Equip.* **2017**, *31*, 499–504. [[CrossRef](#)]
49. Thevenot, M.; Dignac, M.F.; Rumpel, C. Fate of lignins in soils: A review. *Soil Biol. Biochem.* **2010**, *42*, 1200–1211. [[CrossRef](#)]
50. Dümig, A.; Knicker, H.; Schad, P.; Rumpel, C.; Dignac, M.F.; Kögel-Knabner, I. Changes in soil organic matter composition are associated with forest encroachment into grassland with long-term fire history. *Eur. J. Soil Sci.* **2009**, *60*, 578–589. [[CrossRef](#)]
51. Osmaston, H.A.; Mitchell, W.A.; Osmaston, J.A.N. Quaternary glaciation of the Bale Mountains, Ethiopia. *J. Quat. Sci.* **2005**, *20*, 593–606. [[CrossRef](#)]
52. Billi, P. Geomorphological landscapes of Ethiopia. In *Landscapes and Landforms of Ethiopia*; Springer: Dordrecht, The Netherlands, 2015; pp. 3–32.
53. Woldu, Z.; Feoli, E.; Nigatu, L. Partitioning an Elevation Gradient of Vegetation from Southeastern Ethiopia by Probabilistic Methods. *Plateau* **1989**, *81*, 189–198.
54. Bonnefille, R. Evidence for a cooler and drier climate in the Ethiopian uplands towards 2.5 Myr ago. *Nature* **1983**, *303*, 487–491. [[CrossRef](#)]

55. Admasu, E.; Thirgood, S.J.; Bekele, A.; Karen Laurenson, M. Spatial ecology of golden jackal in farmland in the Ethiopian Highlands. *Afr. J. Ecol.* **2004**, *42*, 144–152. [[CrossRef](#)]
56. Kebede, S.; Travi, Y. Origin of the $\delta^{18}\text{O}$ and $\delta^2\text{H}$ composition of meteoric waters in Ethiopia. *Quat. Int.* **2012**, *257*, 4–12. [[CrossRef](#)]
57. Segele, Z.T.; Lamb, P.J.; Leslie, L.M. Large-scale atmospheric circulation and global sea surface temperature associations with horn of Africa June–September rainfall. *Int. J. Climatol.* **2009**, *29*, 1075–1100. [[CrossRef](#)]
58. Dullo, B.W.; Grootjans, A.P.; Roelofs, J.G.M.; Senbeta, A.F.; Fritz, C. Fen mires with cushion plants in Bale Mountains, Ethiopia. *Mires Peat* **2015**, *15*, 1–10.
59. Hedges, J.I.; Ertel, J.R. Characterization of Lignin. *Anal. Chem.* **1982**, *54*, 174–178. [[CrossRef](#)]
60. Goñi, M.A.; Hedges, J.I. Lignin dimers: Structures, distribution, and potential geochemical applications. *Geochim. Cosmochim. Acta* **1992**, *56*, 4025–4043. [[CrossRef](#)]
61. Amelung, W.; Martius, C.; Bandeira, A.G.; Garcia, M.V.B.; Zech, W. Lignin characteristics and density fractions of termite nests in an Amazonian rain forest—Indicators of termite feeding guilds? *Soil Biol. Biochem.* **2002**, *34*, 367–372. [[CrossRef](#)]
62. Möller, A.; Kaiser, K.; Zech, W. Lignin, carbohydrate, and amino sugar distribution and transformation in the tropical highland soils of northern Thailand under cabbage cultivation, Pinus reforestation, secondary forest, and primary forest. *Aust. J. Soil Res.* **2002**, *40*, 977–998. [[CrossRef](#)]
63. R Core Team. R: A Language and Environment for Statistical Computing. R Foundation for Statistical Computing: Vienna, Austria, 2019. Available online: <http://www.R-project.org/> (accessed on 26 April 2019).



© 2019 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<http://creativecommons.org/licenses/by/4.0/>).