New Phytologist

Research

Signals of speciation: volatile organic compounds resolve closely related sagebrush taxa, suggesting their importance in evolution

Deidre M. Jaeger[1], Justin B. Runyon[2] and Bryce A. Richardson[1]

[1]USDA Forest Service, Rocky Mountain Research Station, 735 N. 500 East, Provo, UT 84606, USA; [2]USDA Forest Service, Rocky Mountain Research Station, 1648 S. 7th Avenue, Bozeman, MT 59717, USA

Author for correspondence:
Bryce A. Richardson
Tel: +1 801 356 5112
Email:brichardson02@fs.fed.us

Received: *5 January 2016*
Accepted: *22 March 2016*

New Phytologist (2016) **211:** 1393–1401
doi: 10.1111/nph.13982

Key words: *Artemisia tridentata*, common garden, dynamic headspace, GC-MS, plant volatiles, Random Forest.

Summary

• Volatile organic compounds (VOCs) play important roles in the environmental adaptation and fitness of plants. Comparison of the qualitative and quantitative differences in VOCs among closely related taxa and assessing the effects of environment on their emissions are important steps to deducing VOC function and evolutionary importance.

• Headspace VOCs from five taxa of sagebrush (*Artemisia*, subgenus *Tridentatae*) growing in two common gardens were collected and analyzed using GC-MS.

• Of the 74 total VOCs emitted, only 15 were needed to segregate sagebrush taxa using Random Forest analysis with a low error of 4%. All but one of these 15 VOCs showed qualitative differences among taxa. Ordination of results showed strong clustering that reflects taxonomic classification. Random Forest identified five VOCs that classify based on environment (2% error), which do not overlap with the 15 VOCs that segregated taxa.

• We show that VOCs can discriminate closely related species and subspecies of *Artemisia*, which are difficult to define using molecular markers or morphology. Thus, it appears that changes in VOCs either lead the way or follow closely behind speciation in this group. Future research should explore the functions of VOCs, which could provide further insights into the evolution of sagebrushes.

Introduction

Plants synthesize and emit a variety of volatile organic compounds (VOCs) that serve diverse physiological and ecological functions. Plant VOCs provide protection against biotic and abiotic stresses (Loreto & Schnitzler, 2010) and mediate a large array of interactions with other organisms. For example, VOCs convey information about plant identity, status, and location to beneficial organisms like pollinators (Raguso, 2008; Schiestl, 2015) as well as to plant enemies like insect herbivores (Bruce *et al.*, 2005; Gray *et al.*, 2015) and parasitic plants (Runyon *et al.*, 2006). Plant VOCs also play important roles in defense directly by repelling insect herbivores (De Moraes *et al.*, 2001) or inhibiting bacterial growth (Huang *et al.*, 2012) and indirectly by attracting natural enemies of herbivores (Schuman *et al.*, 2012; Amo *et al.*, 2013). Herbivore-induced VOCs can also function within plants to signal imminent attack to distant plant parts (Heil & Bueno, 2007), and neighboring plants can eavesdrop on these signals to anticipate attack (Karban *et al.*, 2014b). Given these important roles, VOCs are probably under strong selection and shaped by these environmental interactions (Dicke & Baldwin, 2010).

Members of the *Artemisia* subgenus *Tridentatae* (Asteraceae) are highly aromatic woody shrubs known as sagebrushes. They are believed to be a relatively young species group in North America. Eurasian progenitors to this subgenus probably colonized North America via Beringia and radiated into a variety of species and subspecies during the Pliocene and Pleistocene (McArthur & Sanderson, 1999). Currently, 15 species and 12 subspecies of sagebrush are recognized (Shultz, 2009). Among sagebrushes, *Artemisia tridentata* (big sagebrush) is the most abundant and widespread species. Three of the five subspecies of big sagebrush are common and dominant across arid western North America. Each of these subspecies is adapted to a particular environmental niche. Mountain big sagebrush (*Artemisia tridentata* ssp. *vaseyana*) grows on mountain slopes at higher elevations, basin big sagebrush (*Artemisia tridentata* ssp. *tridentata*) is found in drainage basins with deep soils, and Wyoming big sagebrush (*Artemisia tridentata* ssp. *wyomingensis*) also grows in basins but in shallow, drier soils. Two closely related species, *Artemisia arbuscula* (low sagebrush) and *Artemisia nova* (black sagebrush), can grow in close proximity to *A. tridentata* but typically occupy different soil types (Mahalovich & McArthur, 2004). In each of these niches, sagebrush plants are probably exposed to distinct abiotic stresses and interactions with herbivores or competitors that could differentially influence the evolution of VOCs.

Taxonomic treatments using morphology and molecular genetic markers have had difficulties in discerning relationships within the *Tridentatae*. Among the taxonomic treatments, there have been several revisions (reviewed in Shultz, 2009) owing to variability in morphological traits. Moreover, molecular genetic

studies have failed to clearly resolve phylogenetic relationships among some species of this subgenus (Garcia *et al.*, 2011). Richardson *et al.* (2012) examined the intraspecific relationships of *A. tridentata* using secondary metabolite genes, finding that only a few of the 24 genes carried a phylogenetic signal supporting clades of *A.t. tridentata* and *A.t. vaseyana*. The exclusively tetraploid *A.t. wyomingensis* was polyphyletic among *A.t. tridentata* and *A.t. vaseyana* clades (Richardson *et al.*, 2012). This study and others suggest that a combination of relatively recent speciation events, as well as past and current hybridization among closely related species and subspecies, have probably confounded phylogenetic and taxonomic interpretation of the *Tridentatae*.

Despite these taxonomic uncertainties in the *Tridentatae*, plant chemistry can be used to separate some species and subspecies. For example, large amounts of within-leaf coumarins, which fluoresce under ultraviolet light, are indicative of *A.t. vaseyana* as it is the only *A. tridentata* subspecies that produces coumarins in abundance (Stevens & McArthur, 1974; Shumar *et al.*, 1982; McArthur *et al.*, 1988). Earlier studies of within-leaf monoterpenes and phenolics showed significant differences among sagebrush taxa (Welch & McArthur, 1981; Kelsey *et al.*, 1983; Wilt & Miller, 1992), but other studies pointed out considerable seasonal variation in these chemical compounds (Kelsey *et al.*, 1982; Cedarleaf *et al.*, 1983; Wilt & Miller, 1992). These studies show that plant chemistry varies among sagebrush taxa and holds some taxonomic value, but the genetic and environmental influences underlying chemical differences is not known. Likewise, the taxonomic value and genetic and environmental control of sagebrush volatile chemistry have not been studied and there is no reliable way to confidently discern among the three subspecies of big sagebrush.

The diverse habitat niches of sagebrush taxa have probably been predominant factors shaping VOCs through abiotic stress, herbivore interactions, and plant competition. Understanding VOC variation among taxa could be important for restoration of sagebrush ecosystems that are imperiled by disturbance and nonnative weeds. For example, successful restoration of sagebrush habitat will depend on identification of sagebrush taxa, so the right plant can be used in the right place. At present, identification is difficult, particularly for subspecies of *A. tridentata* (Richardson *et al.*, 2015). Moreover, many native wildlife species, like the threatened greater sage-grouse (*Centrocercus urophasianus*) and pygmy rabbit (*Brachylagus idahoensis*), are dependent on sagebrush and have a preference for foraging on specific taxa, which is thought to be guided by VOCs (Frye *et al.*, 2013; Ulappa *et al.*, 2014). In this study, we collected sagebrush VOCs in common gardens to: determine if VOCs can be used to distinguish sagebrush taxa; identify the VOCs most important for classification; and assess which VOCs have small and large environmental effects.

Materials and Methods

Common gardens and plants sampled

Volatile organic compounds were sampled from three species of *Artemisia*: *A. arbuscula* Nutt., *A. nova* Nutt., and *A. tridentata* A Nelson. To focus on intraspecific variation within *A. tridentata*, three subspecies were included: *tridentata*, *vaseyana*, and *wyomingensis*. VOC samples from two big sagebrush hybrids (*A.t. tridentata* × *A.t. vaseyana*) were also included for additional comparisons. The majority (94%) of VOCs were collected from plants growing in two common gardens. Seeds from 10 wild populations of *A. tridentata* and one wild population of *A. arbuscula* were collected in autumn of 2009 from locations in California, Oregon, Idaho and Utah (Fig. 1; Supporting Information Table S1). Seedlings were grown in a glasshouse for 3 months, hardened outside for 2 wk, and planted into the common gardens in the spring of 2010. Seedlings derived from a single wild-collected plant are maternal half-sibs, hereafter referred to as a family.

The two common gardens were stationed in contrasting climates. Majors Flat garden is located near Ephraim, UT, USA (39.339°N, 111.578°W; 2205 m elevation) and Orchard garden is located near Boise, ID, USA (43.322°N, 115.998°W; 974 m elevation) (Fig. 1; Richardson *et al.*, 2015). Majors Flat is a relatively cooler and mesic canyon where *A.t. vaseyana* co-occurs with Gambel oak (*Quercus gambelii* Nutt.) and Utah juniper (*Juniperus osteosperma* (Torr.) Little). The mean annual temperature at Majors Flat is 8°C and the annual precipitation is 414 mm. The Orchard garden is in a warm and dry basin dominated by *A.t. wyomingensis* with *A.t. tridentata* occurring in deeper soils. Mean annual temperature at Orchard is 10.9°C and annual

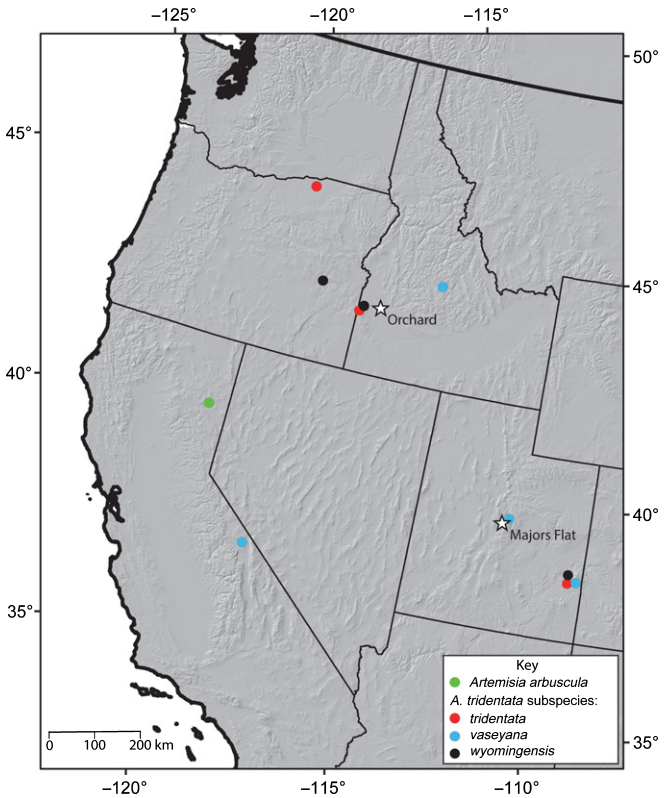

Fig. 1 Map of the western USA showing Majors Flat, UT, and Orchard, ID, common gardens and source population locations for *Artemisia arbuscula* and the three subspecies of *Artemisia tridentata* – *tridentata*, *vaseyana*, *wyomingensis* – used in this study.

precipitation is 224 mm. These gardens have a completely randomized block design and are enclosed by a wire fence that is 6 feet (1.83 m) tall above ground and continues 2 feet (0.61 m) below ground to discourage herbivory.

Artemisia nova plants and big sagebrush hybrids (*A.t. tridentata* × *A.t. vaseyana*) growing at the Shrub Science Lab in Provo, UT, USA, were sampled to further examine variation in VOCs. *Artemisia nova* seeds were originally collected from a population near Yuba Lake, UT (39.4016°N, 112.0312°W), in 2011. The seeds were germinated in a glasshouse and planted in a plot in 2012. For the hybrid-controlled cross, seeds were collected from experimental plots in Hobble Creek Canyon, UT, as described by McArthur *et al.* (1988) and Weber *et al.*(1994). These hybrid plants were F_2 plants from seed of F_1 plants. The parents of the experimental F_1 hybrids were originally from wild populations. *A.t. vaseyana* (male parent) was from Hobble Creek, near Springville, UT, and *A.t. tridentata* (female parent) was collected near Dove Creek, CO, and transplanted as seedlings to the Snow Field Station, Ephraim, UT, USA. Experimental hybridizations were performed in 1980 at the Snow Field Station (McArthur *et al.*, 1988). Subsequently, seeds from the F_1 plants were grown to maturity at the Snow Field Station and seeds from those plants (F_2s) were collected and grown in experimental plots in Hobble Creek Canyon (Weber *et al.*, 1994). Some of the F_2 hybrid seeds were stored at the Shrub Sciences Laboratory and grown in 1997, remaining in the glasshouse for 15 yr until outplanted to the Shrub Science Lab garden in 2012.

Subspecies and cytotype of the samples were determined by a combination of morphology, UV fluorescence, flow cytometry and genetic markers. As described in Richardson *et al.* (2012), tissue samples were used to evaluate ploidy and UV fluorescence (Stevens & McArthur, 1974) for all samples from the Shrub Sciences Laboratory and Orchard garden. Select plants ($n = 12$) at Majors Flat were also tested for UV fluorescence and ploidy using flow cytometry. These plants were examined because VOC ordination revealed two collections that diverged from the expected taxonomic classifications (Table S1).

Volatile collection and analysis

Volatile organic compounds from the common gardens were collected on 29 July and 2 October 2014 at Majors Flat and on 20 May 2015 at Orchard. On each collection date, 30–34 individuals were sampled, representing 10–11 populations. The sampling strategy for each collection date was to replicate sampling of the same populations and half-sib families. However, 14 families, primarily *A.t. vaseyana*, could not be replicated at the Orchard garden as a result of mortality. Volatiles were collected from a total of 11 populations and 43 families. The species and subspecies composition of these populations are shown in Table S1. Four *A. nova* individuals and two hybrids were sampled on July 1, 2015.

Volatiles were collected from plants with no obvious signs of stress (herbivory, pathogen attack, etc.) using portable volatile collection systems comprising automated vacuum pumps (Volatile Assay Systems, Rensselaer, NY, USA) or other small air sampling vacuum pumps (AirLite; SKC Inc., Eighty Four, PA, USA). Approximately 40 cm of an apical branch was enclosed in clear Teflon bags (50 cm wide × 75 cm deep; American Durafilm Co., Holliston, MA, USA) and air was pulled out through a side port (0.5 l min^{-1}) through volatile traps containing 30 mg of the porous polymer adsorbent HayeSep-Q (Restek, Bellefonte, PA, USA). Volatile emissions were collected for 15 or 30 min between 09:00 and 15:00 h and amounts were standardized on a per-h basis for analysis (ng h^{-1}). Volatiles were eluted from traps with 200 µl of dichloromethane, and 1000 ng of *n*-nonyl acetate was added as the internal standard. Samples were analyzed using an Agilent 7890A GC coupled with a 5975C MS and separated on an HP-1 ms (30 m × 0.25 mm inside diameter, 0.25 µm film thickness) column; helium was used as the carrier gas. The GC oven was maintained at 35°C for 3 min and then increased by 4°C min^{-1} to 175°C, then 25°C min^{-1} to 250°C.

Quantifications were made relative to the internal standard using ChemStation software (Agilent Technologies, Wilmington, DE, USA). Identification of compounds were made using the NIST 08 Mass Spectral Search Program (National Institute of Standards and Technology, Gaithersburg, MD, USA) and confirmed by comparing retention times and mass spectra with commercial standards, when available. If commercial standards were not available, compounds were named if matched using NIST 08 Mass Spectral Search Program and previously reported in the literature to occur in the genus *Artemisia*. The remaining unidentified compounds were labeled as unidentified monoterpenoids (MT1, MT2, etc.) or unidentified sesquiterpenoids (ST1, ST2, etc.) and these compounds are also presented in Table S2. Methacrolein, a four-carbon VOC previously reported to occur in sagebrush, could not be examined with this methodology (hidden by solvent peak).

Statistical analysis

We used the Random Forest classification algorithm (Breiman, 2001) to investigate whether VOCs hold value for distinguishing *Artemisia* species and subspecies. Random Forest classification is increasingly being used to analyze large and complex data sets, including plant VOCs (Ranganathan & Borges, 2010). Two Random Forest analyses were conducted to assess the importance of VOCs using two classification schemes: genetic (i.e. subspecies/species groups) and environmental effects (i.e. garden and collection date). Because of the interaction between garden and collection date, the two variables were grouped into an interaction term referred to as environment. These analyses used randomForest package v.4.6-10 (Liaw & Wiener, 2002) and varSELRF v.4.6-10 (Díaz-Uriarte & Alvarez de Andrés, 2006) in R v.3.1.2 (R Core Team, 2014). The package varSELRF was used to select VOC models that best classified each scheme, genetic and environment, with the smallest out-of-bag error. Variable elimination was carried out using out-of-bag error with ntrees set at 1000, and model selection was based on 200 bootstrap replicates. All other parameters were set as the defaults. VOC importance for each classification scheme was ranked with mean decrease in accuracy (MDA). To visualize classification

based on genetics and the environment, multidimension scaling plots were used, part of the randomForest package.

Total VOC production among the common garden samples was evaluated quantitatively and qualitatively using a linear mixed-model, package LME4 in R v.3.1.2 (R Core Team, 2014), with independent variables divided into fixed and random effects. Environments (garden and collection date) were treated as random effects and genetics (species/subspecies) as fixed effects. Total VOCs for *A. nova* and the hybrid were not included because of considerably lower sample size than the common garden groups.

Results

A total of 74 VOCs were detected using GC-MS analysis of 100 headspace collections from 70 plants representing five sagebrush taxa (Tables S1, S2). Monoterpenoids (62 compounds; 84%) dominated the composition of VOCs, followed by sesquiterpenoids (eight compounds; 11%), green leaf volatiles (two compounds; 2%) and heterocyclic compounds (two compounds; 2%). We were able to identify 58 (78%) of the compounds (Table S2). We confirmed 26 VOCs by comparison with commercial standards and another 32 VOCs were putatively identified using NIST software together with published literature on *Artemisia* volatiles (Table S2). There were 16 VOC compounds that could not be fully identified: 12 compounds were labeled as unknown monoterpenoids (MT) and four were labeled as unknown sesquiterpenoids (ST).

Overall, VOC qualitative differences among taxa were more prevalent than quantitative differences. The total amounts of VOCs emitted (ng h^{-1}) did not differ among taxa, common garden, or collection date except for *A. arbuscula*, which emitted significantly more total VOCs than big sagebrush subspecies ($P = 0.0308$). Only 10 volatiles were emitted by all plants on all collection dates (Table S2). Some VOCs were unique to, and only emitted by, one taxon. For example, artemisia alcohol 2 and *trans*-verbenol were produced solely by *A. arbuscula*, whereas 3-carene, *cis*-geraniol, beta-myrcene, nerol acetate, and alpha-terpinene were only emitted by *A. nova*. *Cis*-verbenol was emitted by *A. nova* and *A. arbuscula* but not *A. tridentata*. Six compounds were emitted only by *A. tridentata*: *p*-menth-1-en-9-al, *trans*-hotrienol, 5,-5-dimethyl-2-(5H)-Furanone, and MT2, MT3, MT4. Among *A. tridentata* subspecies, *A.t. tridentata* was the only subspecies that did not emit artemisia ketone and *cis*-sabinene. *A.t. wyomingensis* was the only subspecies that did not emit MT7, but almost exclusively emitted artemisia ketone 2, 5,5-dimethyl-2-(5H)-Furanone, MT2, MT3 and MT4. *A.t. vaseyana* was the only subspecies that did not emit artemisia triene and the only one to emit alpha- and beta-thujone, 4-terpineol, and *trans*-myrtenyl acetate.

Similar to Karban *et al.* (2014a), we identified thujone and camphor chemotypes in which plants from the same population emitted either large amounts of alpha- and beta-thujone and low amounts of camphor or large amounts of camphor and no thujone. The thujone/camphor chemotypes appear unique to *A.t. vaseyana* as no alpha- or beta-thujone was detected in any *A.t. tridentata* and *A.t. wyomingensis* VOC collections. Camphor was emitted by all plants.

Environment affected emissions of some VOCs. The Random Forest algorithm identified five VOCs that best classified quantitative differences between the gardens and collection dates: two green-leaf volatiles (GLVs), two sesquiterpenoids, and one monoterpene (Table 1). Multidimensional scaling (MDS) with these five VOCs showed strong clustering by garden and collection date (Figs 2, S1). The out-of-bag error for classifying by environment was 2%.

We used the Random Forest algorithm to identify the best discriminator VOCs that could potentially be used for taxonomic classification. Analyses determined that 15 volatiles with the highest mean decrease in accuracy provided a model with the lowest out-of-bag error rate at 4% (Table 2; Fig. S2). Most of these VOCs were monoterpenoids (13 of 15 VOCs). MDS of these 15 volatiles shows strong clustering by species and subspecies (Fig. 3a,b) with dimensions 1 and 2 segregating the subspecies of *A. tridentata* (Fig. 3a) and dimensions 3 and 4 segregating the species *A. tridentata*, *A. arbuscula*, and *A. nova* (Fig. 3b). These compounds displayed predominantly qualitative differences (14 of 15 VOCs), in which each compound was emitted by one to four of the five species/subspecies and absent from the other taxa (Table 2). None of the five VOCs used in the environmental Random Forest classification overlapped with the VOCs identified by Random Forest in the genetically determined VOC model (Tables 1, 2).

Random forest classification identified four outliers (three plants) that did not match our taxonomic identification. One of these plants was an *A.t. tridentata* × *vaseyana* hybrid, which the model classified as *A.t. vaseyana*. Two were suspected *A.t. wyomingensis*, ORW1-1, collected at Majors Flat in July and October, which VOCs classified as *A.t. tridentata*. VOCs from ORW1-1 were also collected from Orchard; however, the VOC profile does not resemble the VOC profiles of the ORW1-1 collected at Majors Flat, indicating that the plants are different (see the Discussion section). The final outlier was one *A.t. vaseyana* that classified as *A. arbuscula*.

Table 1 Importance ranking of volatile organic compounds (VOCs) in classifying 100 *Artemisia* collections based on environment (garden and collection date)

Importance rank	VOCs	MDA
1	ST2	57.13
2	*cis*-3-hexenyl acetate[1]	38.21
3	*cis*-3-hexenol[1]	33.84
4	germacrene-D[2]	24.56
5	D-limonene[1]	19.53

Rankings are based on mean decreasing accuracy (MDA) using Random Forest.
ST, unidentified sesquiterpenoid.
[1]Identity verified by comparing retention time and mass spectrum with authentic standard.
[2]Identified using NIST 08 Mass Spectral Search Program and comparison with published literature on *Artemisia* VOCs.

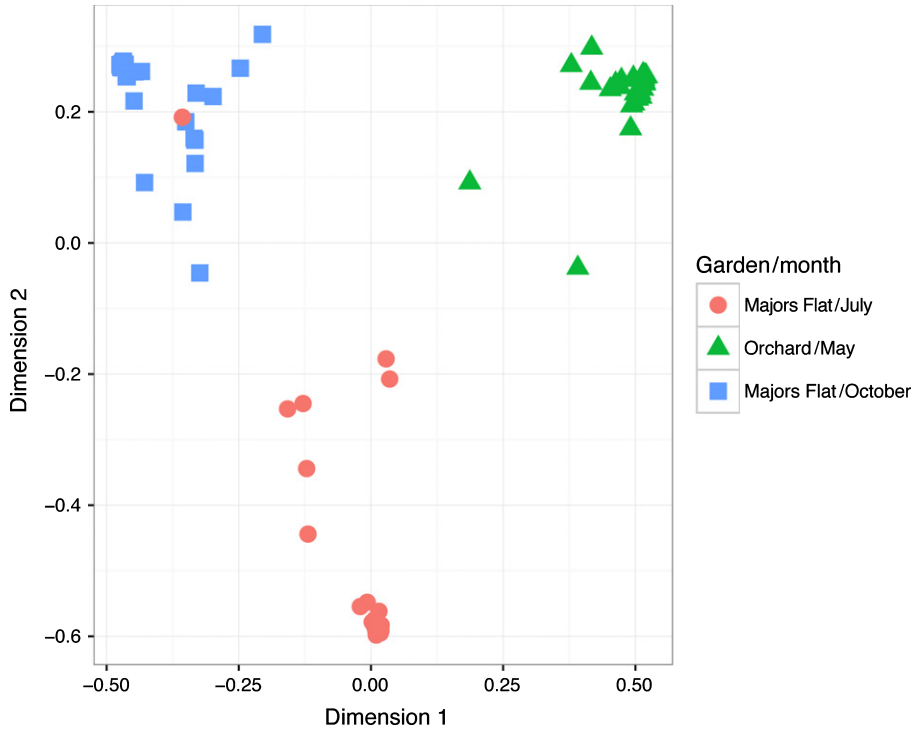

Fig. 2 Sagebrush volatiles can discern environmental effects. Nonmetric multidimensional scaling plot of Random Forest model classifying environment (interaction of garden and collection date) using five volatile compounds (Table 1). Species and subspecies included are: *Artemisia arbuscula*, hybrid = *Artemisia tridentata tridentata* × *Artemisia tridentata vaseyana*, *Artemisia nova*, *A.t. tridentata*, *A.t. vaseyana*, *A.t. wyomingensis*. The out-of-bag error is 2%.

Table 2 Importance ranking and distribution of volatile organic compounds (VOCs) found among five taxonomic groups and a control hybrid of *Artemisia* using Random Forest classification

Rank (MDA)	VOCs	*Artemisia arbuscula* (12)	*Artemisia tridentata tridentata* (28)	*Artemisia tridentata vaseyana* (26)	*Artemisia tridentata wyomingensis* (28)	*Artemisia nova* (4)	Hybrid (2)
1 (20.6)	*p*-menth-1-en-9-al[1]	0	0	0	26	0	2
2 (19.9)	*cis*-sabinol[1]	12	0	26	2	4	0
3 (15.6)	lavandulol acetate[1]	4	28	2	27	0	0
4 (15.0)	5.5-dimethyl-2-(5H)-Furanone[1]	0	0	0	23	0	0
5 (14.9)	myrtenol[2]	1	25	26	28	0	2
6 (14.8)	4-terpineol[2]	0	0	23	0	4	2
7 (14.0)	ST4	12	28	26	26	0	2
8 (13.9)	*cis*-beta-ocimene[2]	12	28	26	7	4	0
9 (13.6)	*trans*-methyl-santolinate[1]	2	28	15	25	0	0
10 (13.4)	MT near artemisia alcohol[1]	12	28	13	28	0	0
11 (12.4)	sabinene[2]	4	28	26	28	4	2
12 (12.3)	*trans*-beta-santolina-epoxide[1]	12	28	23	28	0	2
13 (12.2)	artemisia alcohol 1[1]	11	28	17	28	0	2
14 (11.3)	*cis*-methyl-santolinate[1]	10	28	26	28	0	0
15 (11.2)	MT2	0	0	0	19	0	0

The total number of collections for each taxon is shown in parentheses. Ranking based on mean decrease in accuracy (MDA). Hybrid = *A. tridentata vaseyana* × *A.t. tridentata* control cross.
MT, unidentified monoterpenoid; ST, unidentified sesquiterpenoid.
[1]Identified using the NIST 08 Mass Spectral Search Program and comparison with published literature on *Artemisia* VOCs.
[2]Identity verified by comparing retention time and mass spectrum with authentic standard.

Discussion

We have shown that VOCs can be used to reliably distinguish a set of closely related species and subspecies of sagebrushes. VOCs appear to be at least as reliable as the current methodology: morphology and within-plant chemistry. VOCs also appear to be more reliable than recent studies employing molecular markers (Richardson *et al.*, 2012). The Random Forest algorithm correctly identified 67 of the 70 (96%) plants sampled using VOCs. Of the 74 total volatile compounds emitted by sagebrushes, only 15 compounds were needed to discriminate the five sagebrush taxa (Table 2). Interestingly, some of these diagnostic compounds are uncommon VOCs that have only been reported from *Artemisia* or closely related taxa (e.g. artemisia alcohol,

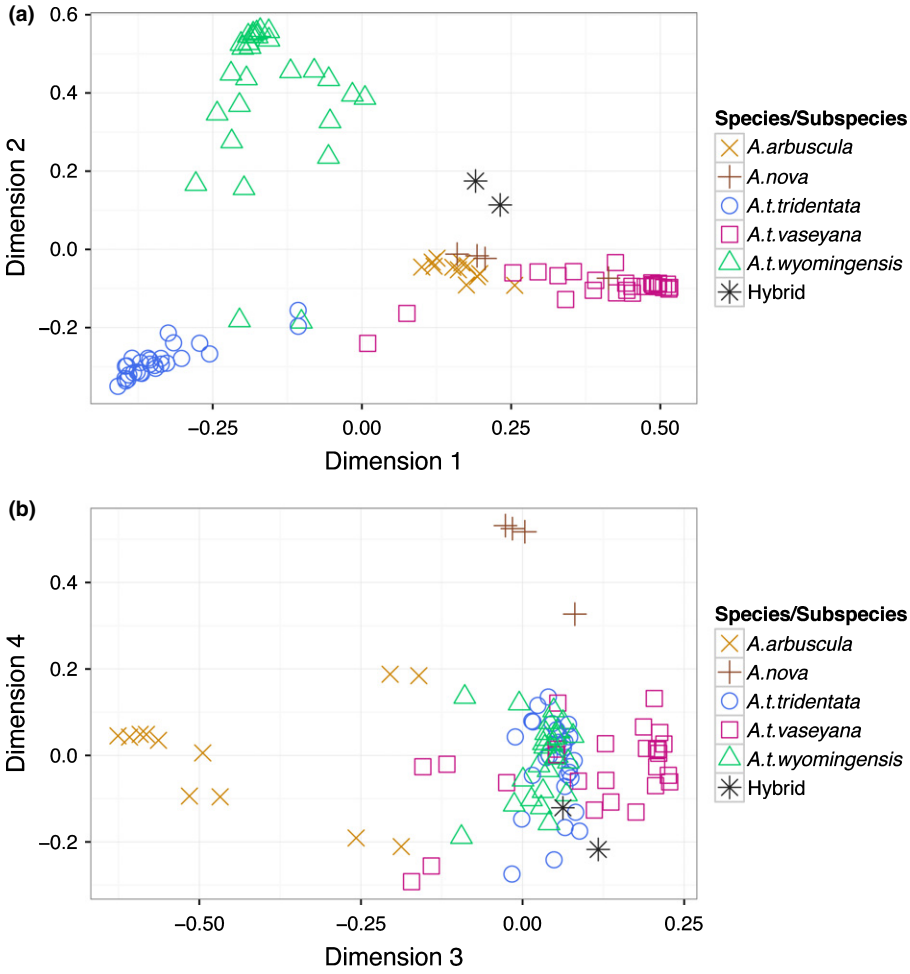

Fig. 3 Plant volatiles can distinguish closely related sagebrush taxa. Nonmetric multidimensional scaling plot of Random Forest model classifying species and subspecies using 15 volatile compounds (Table 2). (a) Dimensions 1 and 2 segregate subspecies of *Artemisia tridentata*; (b) dimensions 3 and 4 segregate species *A. tridentata*, *Artemisia arbuscula* and *Artemisia nova*. The out-of-bag error is 4%.

methyl santolinate, santolina epoxide) (Turi *et al.*, 2014). Fourteen of these 15 VOCs showed qualitative differences among taxa and these differences persisted regardless of environment (i.e. common garden or collection date), suggesting that emission of these compounds is largely under genetic control. Moreover, VOCs most important in classification of taxonomic groups (Table 2) did not overlap with the VOCs most important in separating environments (Table 1). These findings indicate that sagebrush volatile profiles can robustly distinguish species and subspecies identity and could provide a means to identify plants for ecological restoration.

Not surprisingly, environment also affected VOC emissions. Only five volatile compounds were needed to distinguish among interacting environmental components: gardens and sampling dates (Fig. 2; Table 1). Four of the five VOCs important in classifying environment were GLVs (*cis*-3-hexenyl acetate and *cis*-3-hexenol) and sesquiterpenes (germacrene-D and one unidentified sesquiterpene). Both classes of compounds are known to be highly responsive to the environment. Undamaged plants typically produce trace amounts of GLVs and sesquiterpenes, whereas large amounts are emitted in response to various biotic or abiotic stresses (Loreto & Schnitzler, 2010). Herbivore feeding, pathogen infection, high light and temperature, and ozone are each known to induce plants to emit GLVs and/or sesquiterpenes (Gouinguené & Turlings, 2002; Beauchamp *et al.*, 2005; Bourtsoukidis *et al.*, 2012). Moreover, clipping of *A.t. tridentata* foliage is known to trigger emission of GLVs and sesquiterpenes (Kessler *et al.*, 2006). By contrast, no GLVs and only one sesquiterpene are among the most important 15 compounds separating sagebrush taxa (Table 2). This suggests that biotic and/or abiotic differences between gardens and sampling dates are affecting emission of these five VOCs in similar ways for all sagebrush species and subspecies.

Volatile organic compounds from three plants were misidentified by the Random Forest algorithm and did not plot according to expectations. Two VOC collections in July and October from one plant (ORW1-1) in the Majors Flat garden plotted with *A.t. tridentata* (Fig. 3a), but this was identified as *A.t. wyomingensis* based on ploidy and UV fluorescence from a family member at the Orchard garden. Reassessment of cytotype, UV fluorescence, and morphology confirmed this plant to be tetraploid with no UV fluorescence, typical of *A.t. wyomingensis*. However, ORW1-1 exhibited height and growth form more typical of *A.t. tridentata* (B. A. Richardson, unpublished), suggesting this plant could be a tetraploid *A.t. tridentata*, which has been infrequently identified in other studies (McArthur & Sanderson, 1999; Richardson *et al.*, 2012). The second misidentified plant was one *A.t. vaseyana* that classified as *A. arbuscula*. The

A.t. vaseyana VOC cluster is closest in proximity to *A. arbuscula* in the MDS ordination (Fig. 3b), so perhaps hybrid characters can be found between these two taxa; however, further study is necessary to assess interspecific hybridization. And finally, one of two hybrid plants (*A.t. tridentata* × *vaseyana*) classified as *A.t. vaseyana*. The paternal parent is *A.t. vaseyana*, so it is not a surprising classification; however, the other hybrid plant from this controlled cross was classified separately. UV fluorescence, ploidy, and morphology are identical for these plants, and both hybrids clustered in the same general area (Fig. 3a). Additional sampling of hybrid plants is needed and could refine the set of VOCs used by Random Forest to characterize sagebrush taxa and improve recognition of hybrids.

Based on the complex nature of sagebrush taxonomy, the variability of morphological traits within taxa, and occasional shared traits among taxa, the accuracy to which VOCs are able to discern these closely related species and subspecies is surprising. Compared with other diagnostic techniques (e.g. morphology and DNA), VOCs performed better in distinguishing taxa, especially among subspecies of *A. tridentata*. Even controlled hybrids appear to behave as expected with intermediate MDS coordinates between hybrid progenitors (Fig. 3a). For the sagebrushes, both hybridization and relatively recent speciation have been suspected in the variable morphological characteristics observed within taxa and shared characters among taxa (McArthur & Sanderson, 1999; Shultz, 2009) and the discordance in resolving phylogenies (Richardson *et al.*, 2012). This study and previous investigations (Stevens & McArthur, 1974; Kelsey *et al.*, 1983) suggest plant chemistry could be a more informative set of characteristics for deducing taxonomy in sagebrushes than other morphological and DNA diagnostics. Moreover, it suggests VOCs may be key in environmental adaptation and diverged during speciation.

There are many factors that could explain why plant VOCs resonate with the taxonomy of sagebrushes. VOCs could have adapted to, or aided expansion into, the different habitats in which each species and subspecies occurs. For example, some sagebrush VOCs are allelopathic (Preston *et al.*, 2002) and VOCs could be tailored to the specific plant competitors encountered in each habitat (Davies & Bates, 2010). VOCs, especially monoterpenes, can protect plants from oxidative and thermal stresses (Loreto & Schnitzler, 2010), which could allow plants to grow in more xeric environments. Herbivores and pathogens are probably important drivers of VOC patterns in sagebrushes. Sagebrushes are host plants for many specialist and generalist herbivores that can have severe negative effects on plant growth and reproduction (Takahashi & Huntly, 2010). VOCs are undoubtedly serving important roles in mediating interactions with, and defending against, herbivores, as they do for other plant species (Gershenzon & Dudareva, 2007; Schuman *et al.*, 2012; Frye *et al.*, 2013; Ulappa *et al.*, 2014). Similarly, many pathogens infect sagebrushes and can do so in habitat-specific ways (Talley *et al.*, 2002). For example, snow molds primarily affect mountain big sagebrush (*A.t. vaseyana*) which grows at high elevations with deep snow cover (Nelson & Sturges, 1986). Many VOCs, including monoterpenes, are antimicrobial (Gershenzon & Dudareva, 2007; Marei *et al.*, 2012) and the identity and blend of monoterpenes emitted by *A.t. vaseyana* could protect against these pathogens. Lastly, VOCs could, in part, be shaped by the roles they play in airborne communication within and between sagebrush plants. Damage-induced sagebrush VOCs stimulate neighboring sagebrush plants to adjust their defenses, which reduces herbivore damage (Karban *et al.*, 2006) and increases receiver survival, growth, and reproduction (Karban *et al.*, 2012). VOCs function similarly within sagebrush plants, allowing communication among branches (Karban *et al.*, 2006). The emission of distinct VOCs could allow recognition of, and preferential response to, signals from the same species or even relatives (Karban *et al.*, 2014a). Regardless of the mechanisms underlying VOC differences among taxa, VOCs are likely to be an important driver of evolution for the sagebrushes.

In summary, we show that analysis of VOCs coupled with a multivariate classification algorithm (Random Forest) can provide accurate identification of closely related sagebrush species and subspecies. This suggests that VOCs are key to the ecological adaptation and evolution of sagebrushes. Also, these findings could allow identification of sagebrushes for restoration, for example by using an electronic nose (Laothawornkitkul *et al.*, 2008) to match the proper plant with the proper habitat (Richardson *et al.*, 2015). Future research should focus on ascertaining what functions VOCs are playing in each habitat for different sagebrushes – with priority given to those compounds identified in Table 2 – the results of which could provide insights into the evolution of this important group of plant species.

Acknowledgements

The authors thank Hector Ortiz, Dr Jennifer Forbey, Marcella Fremgen and Jordan Nobler for collection assistance and the Utah Department of Natural Resources for common garden maintenance. Thanks to Drs Durant McArthur, Casey Delphia, Jennifer Forbey and two anonymous reviewers for thoughtful comments on the manuscript. Funding was provided by the USDA Forest Service climate change funds and the Great Basin Native Plant Program.

Author contributions

J.B.R. and B.A.R. planned and designed the research. D.M.J., J.B.R. and B.A.R. conducted the fieldwork, performed the experiments, analysed the data and wrote the manuscript.

References

Amo L, Jansen JJ, Dam NM, Dicke M, Visser ME. 2013. Birds exploit herbivore-induced plant volatiles to locate herbivorous prey. *Ecology Letters* **16**: 1348–1355.

Beauchamp J, Wisthaler A, Hansel A, Kleist E, Miebach M, Niinemets Ü, Schurr U, Wildt J. 2005. Ozone induced emissions of biogenic VOC from tobacco: relationships between ozone uptake and emission of LOX products. *Plant, Cell & Environment* **28**: 1334–1343.

Bourtsoukidis E, Bonn B, Dittmann A, Hakola H, Hellén H, Jacobi S. 2012. Ozone stress as a driving force of sesquiterpene emissions: a suggested parameterisation. *Biogeosciences* **9**: 4337–4352.

Breiman L. 2001. Random forests. *Machine Learning* **45**: 5–32.

Bruce TJA, Wadhams LJ, Woodcock CM. 2005. Insect host location: a volatile situation. *Trends in Plant Science* **10**: 269–274.

Cedarleaf JD, Welch BL, Brotherson JD. 1983. Seasonal variation of monoterpenoids in big sagebrush (*Artemisia tridentata*). *Society for Range Management* **36**: 492–494.

Davies KW, Bates JD. 2010. Native perennial forb variation between mountain big sagebrush and Wyoming big sagebrush. *Environmental Management* **46**: 452–458.

De Moraes CM, Mescher MC, Tumlinson JH. 2001. Caterpillar-induced nocturnal plant volatiles repel conspecific females. *Nature* **410**: 577–580.

Díaz-Uriarte R, Alvarez de Andrés S. 2006. Gene selection and classification of microarray data using random forest. *BMC Bioinformatics* **7**: 3.

Dicke M, Baldwin IT. 2010. The evolutionary context for herbivore-induced plant volatiles: beyond the 'cry for help'. *Trends in Plant Science* **15**: 167–175.

Frye GG, Connelly JW, Musil DD, Forbey JS. 2013. Phytochemistry predicts habitat selection by an avian herbivore at multiple spatial scales. *Ecology* **94**: 308–314.

Garcia S, Garnatje T, McArthur ED, Pellicer J, Sanderson SC, Vallès J. 2011. Taxonomic and nomenclatural rearrangements in *Artemisia* subgen. *Tridentatae*, including a redefinition of *Sphaeromeria* (Asteraceae, Anthemideae). *Western North American Naturalist* **71**: 158–163.

Gershenzon J, Dudareva N. 2007. The function of terpene natural products in the natural world. *Nature Chemical Biology* **3**: 408–414.

Gouinguené SP, Turlings TCJ. 2002. The effects of abiotic factors on induced volatile emissions in corn plants. *Plant Physiology* **129**: 1296–1307.

Gray CA, Runyon JB, Jenkins MJ, Giunta AD. 2015. Mountain pine beetles use volatile cues to locate host limber pine and avoid non-host Great Basin bristlecone pine. *PLoS ONE* **10**: e0135752.

Heil M, Bueno JCS. 2007. Herbivore-induced volatiles as rapid signals in systemic plant responses: how to quickly move the information? *Plant Signaling & Behavior* **2**: 191–193.

Huang M, Sanchez-Moreiras AM, Abel C, Sohrabi R, Lee S, Gershenzon J, Tholl D. 2012. The major volatile organic compound emitted from *Arabidopsis thaliana* flowers, the sesquiterpene (E)-caryophyllene, is a defense against a bacterial pathogen. *New Phytologist* **193**: 997–1008.

Karban R, Ishizaki S, Shiojiri K. 2012. Long-term demographic consequences of eavesdropping for sagebrush. *Journal of Ecology* **100**: 932–938.

Karban R, Shiojiri K, Huntzinger M, McCall AC. 2006. Damage-induced resistance in sagebrush: volatiles are key to intra- and interplant communication. *Ecology* **87**: 922–930.

Karban R, Wetzel WC, Shiojiri K, Ishizaki S, Ramirez SR, Blande JD. 2014a. Deciphering the language of plant communication : volatile chemotypes of sagebrush. *New Phytologist* **204**: 380–385.

Karban R, Yang LH, Edwards KF. 2014b. Volatile communication between plants that affects herbivory: a meta-analysis. *Ecology Letters* **17**: 44–52.

Kelsey RG, Stephens JR, Shafizadeh F. 1982. The chemical constituents of sagebrush foliage and their isolation. *Journal of Range Management* **35**: 617–622.

Kelsey RG, Wright WE, Sneva F, Winward A, Britton C. 1983. The concentration and composition of big sagebrush essential oils from Oregon. *Biochemical Systematics and Ecology* **11**: 353–360.

Kessler A, Halitschke R, Diezel C, Baldwin IT. 2006. Priming of plant defense responses in nature by airborne signaling between *Artemisia tridentata* and *Nicotiana attenuata*. *Oecologia* **148**: 280–292.

Laothawornkitkul J, Moore JP, Taylor JE, Possell M, Gibson TD, Hewitt CN, Paul ND. 2008. Discrimination of plant volatile signatures by an electronic nose: a potential technology for plant pest and disease monitoring. *Environmental Science & Technology* **42**: 8433–8439.

Liaw A, Wiener M. 2002. Classification and regression by randomForest. *R News* **2**: 18–22.

Loreto F, Schnitzler J-P. 2010. Abiotic stresses and induced BVOCs. *Trends in Plant Science* **15**: 154–166.

Mahalovich MF, McArthur ED. 2004. Sagebrush (*Artemisia* spp.) seed and plant transfer guidelines. *Native Plants Journal* **5**: 141–148.

Marei GIK, Abdel Rasoul MA, Abdelgaleil SAM. 2012. Comparative antifungal activities and biochemical effects of monoterpenes on plant pathogenic fungi. *Pesticide Biochemistry and Physiology* **103**: 56–61.

McArthur ED, Sanderson SC. 1999. Cytogeography and chromosome evolution of subgenus *Tridentatae* of *Artemisia* (Asteraceae). *American Journal of Botany* **86**: 1754–1775.

McArthur ED, Welch BL, Sanderson SC. 1988. Natural and artificial hybridization between big sagebrush (*Artemisia tridentata*) subspecies. *Journal of Heredity* **79**: 268.

Nelson DL, Sturges DL. 1986. A snowmold disease of mountain big sagebrush. *Phytopathology* **76**: 946–951.

Preston CA, Betts H, Baldwin IT. 2002. Methyl jasmonate as an allelopathic agent: sagebrush inhibits germination of a neighboring tobacco, *Nicotiana attenuata*. *Journal of Chemical Ecology* **28**: 2343–2369.

Raguso RA. 2008. Wake up and smell the roses: the ecology and evolution of floral scent. *Annual Review of Ecology, Evolution, and Systematics* **39**: 549–569.

R Core Team. 2014. *R: a language and environment for statistical computing*. Vienna, Austria: R Foundation for Statistical Computing.

Ranganathan Y, Borges RM. 2010. Reducing the babel in plant volatile communication: using the forest to see the trees. *Plant Biology* **12**: 735–742.

Richardson BA, Ortiz HG, Carlson SL, Jaeger DM, Shaw NL. 2015. Genetic and environmental effects on seed weight in subspecies of big sagebrush : applications for restoration. *Ecosphere* **6**: 201.

Richardson BA, Page JT, Bajgain P, Sanderson SC, Udall JA. 2012. Deep sequencing of amplicons reveals widespread intraspecific hybridization and multiple origins of polyploidy in big sagebrush (*Artemisia tridentata*; Asteraceae). *American Journal of Botany* **99**: 1962–1975.

Runyon JB, Mescher MC, De Moraes CM. 2006. Volatile chemical cues guide host location and host selection by parasitic plants. *Science* **313**: 1964–1967.

Schiestl FP. 2015. Ecology and evolution of floral volatile- mediated information transfer in plants. *New Phytologist* **206**: 571–577.

Schuman MC, Barthel K, Baldwin IT. 2012. Herbivory-induced volatiles function as defenses increasing fitness of the native plant *Nicotiana attenuata* in nature. *eLife* **2012**: e00007.

Shultz LM. 2009. Monograph of *Artemisia* subgenus *Tridentatae* (Asteraceae-Anthemideae). *Systematic Botany Monographs* **89**: 1–131.

Shumar ML, Anderson JE, Reynolds TD. 1982. Identification of subspecies of big sagebrush by ultraviolet spectrophotometry. *Journal of Range Management* **35**: 60–62.

Stevens R, McArthur ED. 1974. A simple field technique for identification of some sagebrush taxa. *Journal of Range Management* **27**: 325–326.

Takahashi M, Huntly N. 2010. Herbivorous insects reduce growth and reproduction of big sagebrush (*Artemisia tridentata*). *Arthropod–Plant Interactions* **4**: 257–266.

Talley SM, Coley PD, Kursar TA. 2002. Antifungal leaf-surface metabolites correlate with fungal abundance in sagebrush populations. *Journal of Chemical Ecology* **28**: 2141–2168.

Turi CE, Shipley PR, Murch SJ. 2014. North American *Artemisia* species from the subgenus *Tridentatae* (Sagebrush): a phytochemical, botanical and pharmacological review. *Phytochemistry* **98**: 9–26.

Ulappa AC, Kelsey RG, Frye GG, Rachlow JL, Shipley LA, Bond L, Pu X, Sorensen Forbey J. 2014. Plant protein and secondary metabolites influence diet selection in a mammalian specialist herbivore. *Journal of Mammalogy* **95**: 834–842.

Weber DJ, Gang DR, Halls SC, Smith BN, McArthur E. 1994. Inheritance of hydrocarbons in subspecific big sagebrush (*Artemisia tridentata*) hybrids. *Biochemical Systematics and Ecology* **22**: 689–697.

Welch BL, McArthur ED. 1981. Variation of monoterpenoid content among subspecies and accessions of *Artemisia tridentata* grown in a uniform garden. *Journal of Range Management* **34**: 380–384.

Wilt EM, Miller GC. 1992. Seasonal variation of coumarin and flavonoid concentrations in persistent leaves of Wyoming big sagebrush (*Artemisia tridentata* ssp. *wyomingensis*: Asteraceae). *Biochemical Systematics and Ecology* **20**: 53–67.

Supporting Information

Additional Supporting Information may be found online in the supporting information tab for this article:

Fig. S1 Out-of-bag error rate vs number of variables (volatile organic compounds) used in the Random Forest classification of environment.

Fig. S2 Out-of-bag error rate vs number of variables (volatile organic compounds) used in the Random Forest classification of taxa.

Table S1 Location, ploidy, and ultraviolet fluorescence information for 100 VOC collections

Table S2 VOCs collected from headspace of *Artemisia* species and subspecies: *A. nova*, *A. arbuscula,* and the three subspecies of *A. tridentata* (*A.t.* ssp. *tridentata*, *A.t.* ssp. *vaseyana*, *A.t.* ssp. *wyomingensis*)

Please note: Wiley Blackwell are not responsible for the content or functionality of any supporting information supplied by the authors. Any queries (other than missing material) should be directed to the *New Phytologist* Central Office.