Discovery of Germacrene A Synthases in *Barnadesia spinosa*: The First Committed Step in Sesquiterpene Lactone Biosynthesis in the Basal Member of the Asteraceae

Trinh-Don Nguyen[a,c,#], Juan A. Faraldos[a,1,#], Maria Vardakou[b], Melissa Salmon[b], Paul E. O'Maille[a,d,2,*], and Dae-Kyun Ro[c,*]

[a]John Innes Centre, Department of Metabolic Biology, Norwich, NR4 7UH, United Kingdom.

[b]University of East Anglia, School of Biological Sciences, Norwich, NR4 7TJ, United Kingdom.

[c]University of Calgary, Department of Biological Sciences, Calgary, T2N 1N4, Canada

[d]Institute of Food Research, Food & Health Programme, Norwich, NR4 7UA, United Kingdom.

Present address:

[1]School of Chemistry, Cardiff University, Main Building, Park Place, Cardiff CF10 3AT, United Kingdom

[2]Biosciences Division, SRI International, 333 Ravenswood Ave, Menlo Park, CA, 94025, USA

Running title: Germacrene A synthases from *Barnadesia spinosa*

[#]Equal contributing authors

*Co-corresponding authors

(Draft |
Accepted for publication in the Biochemical and Biophysical Research Communications
in September 2016)

Abstract

The Andes-endemic Barnadesioideae plant lineage is the oldest surviving and phylogenetically basal subfamily of the Asteraceae (Compositae), a prolific group of flowering plants with world-wide distribution (~25,000 species) marked by a rich diversity of sesquiterpene lactones (STLs). Intriguingly, there is no evidence that members of the Barnadesioideae produce STLs, specialized metabolites thought to have contributed to the adaptive success of the Asteraceae family outside South America. The biosynthesis of STLs requires the intimate expression and functional integration of germacrene A synthase (GAS) and germacrene A oxidase (GAO) activities to sequentially cyclize and oxidize farnesyl diphosphate into the advanced intermediate germacrene A acid leading to diverse STLs. Our previous discovery of GAO activity conserved across all major subfamilies of Asteraceae, including the phylogenetically basal Barnadesioideae, prompted further investigation of the presence of the gateway GAS enzyme in this subfamily. Herein we isolated two terpene synthases (BsGAS1/BsGAS2) from the basal *Barnadesia spinosa* (Barnadesioideae) that displayed robust GAS activity when reconstituted in microbial hosts and characterized *in vitro*. Despite the apparent lack of STLs in the Barnadesioideae, this work unambiguously confirms for the first time the presence of sesquiterpene synthase genes in the basal genera of the Asteraceae. Furthermore, phylogenetic analysis reveals that the two BsGASs fall into two distinct clades of the Asteraceae's GASs, and BsGAS1 clade is only retained in the evolutionary closer Cichorioideae subfamily, implicating BsGAS2 as likely the ancestral base of most GASs found in the STLs-producing lineages outside the Barnadesioideae. Taken together, the presence of potent GAS and GAO activities in *B. spinosa* suggests the enzymatic capacity for STL biosynthesis emerged prior to the subsequent radiation of STL-producing Asteraceae subfamilies.

(Draft |
Accepted for publication in the Biochemical and Biophysical Research Communications
in September 2016)

Key words: Germacrene A synthase, Sesquiterpene lactone, Asteraceae, *Barnadesia spinosa*

1. INTRODUCTION

Sesquiterpene lactones (STLs) comprise a large collection of highly functionalized cyclic C_{15}-isoprenoid plant metabolites routinely used as markers in chemotaxonomic studies of the Asteraceae (Compositae) [1,2], a prolific group of land plants first appearing in South America around 50 million years ago [3,4]. Although simple germacranolide STLs are sporadically found in other angiosperm families (Magnoliaceae, Lauraceae, Cupressaceae, and Apiaceae), only in the Asteraceae have STLs been retained as major secondary metabolic constituents throughout the diversification of the family. Consequently, the biosynthesis, chemical and structural diversity, and varied physiological properties of STLs [5] are often linked to the evolutionary adaptive success of the Asteraceae family outside South America. The Asteraceae stands as one of the largest families of present day plants comprising 13 subfamilies, 17 tribes, 1,600 genera and more than 24,000 species worldwide [6]. Among subfamilies, comprehensive molecular phylogenetic analysis based on chloroplastidic DNA data [7] place the single Barnadesieae tribe of the Barnadesioideae subfamily at the phylogenetic base of the currently known Asteraceae. The current taxonomy (Figure 1A) was deducted [6,8] after the discovery of a unique large (22 Kb) inversion in the chloroplastidic genomes of all other Asteraceae subfamilies but the 9 genera of the Barnadesioideae, which appears to retain a plesiomorphic form of chloroplast DNA found in angiosperm ancestors [9]. The absence of this uniquely modified form of plastidic DNA distinguishes the Barnadesioideae from the rest of the Asteraceae, suggesting all other subfamilies

(Draft |
Accepted for publication in the Biochemical and Biophysical Research Communications
in September 2016)

could indeed have risen from within the resulted evolutionary gap. Notably, in contrast to the worldwide distribution and outstanding secondary metabolite chemistry of STL-producing subfamilies, the ~90 species (9 genera) comprising the Barnadesioideae, the oldest surviving lineage of the Asteraceae, are geographically constrained to the South American Andes. More importantly, only ubiquitous C30 plant triterpenes and/or simple flavonoids (pigments) have been found in the exudates of the ~45 Barnadesioideae species (*Barnadesia, Chuquiraga, Dasyphylum,* and *Schlechendhalia*) chemically surveyed so far [1,10].

The apparent absence of STLs in the Barnadesioideae, along with most other characteristic types of natural products found in all other derived groups, could be taken to indicate the absence of the requisite biosynthetic enzymes in the basal lineage of the Asteraceae. Since much of the rich and complex chemistry of the Asteraceae tribes appears to have emerged after splitting from the basal ancestor, the appearance of the present-day chemotaxonomic STLs most likely constitutes an apomorphy for the non-Barnadesioideae subfamilies. Costunolide is thought to be the central precursor of the vast structural diversity (germacranolides, elemanolides, eudesmanolides, and guainolides) of STLs in the Asteraceae [11]. At the biochemical level, the synthesis of costunolide requires the enzymatic activity of a class I terpene synthase, germacrene A synthase (GAS), which catalyzes the cyclization of farnesyl pyrophosphate (FPP) to germacrene A [12]. This is followed by the sequential cytochrome P450 monooxygenases-catalyzed oxidations of germacrene A at C12 by germacrene A oxidase (GAO) to form germacrene A acid (GAA), and at C6 by costunolide synthase to form 6α-hydroxy GAA, which is spontaneously converted to costunolide [13,14] (Figure 1B).

Intriguingly, the identification and functional expression of a GAO in *Barnadesia spinosa* (BsGAO) [15] suggests at least part of the STL pathway already exists in the phylogenetic base

(Draft |
Accepted for publication in the Biochemical and Biophysical Research Communications
in September 2016)

Barnadesioideae. This observation prompted our further investigation into the presence of GAS in *B. spinosa*. Here we report the discovery of two paralogues of terpene synthase cDNAs from *B. spinosa*. Using metabolically engineered yeast and *in vitro* biochemical assays, we demonstrated the encoded recombinant enzymes were *bona fide* GAS enzymes by direct comparison to the homologues from lettuce (*Lactuca sativa*) and *Artemisia annua*. This result demonstrates that the basal lineage of the Asteraceae encodes two paralogues of GAS which catalyze the first committed step in STL biosynthesis. In light of phylogenetic analysis, this result suggests that GAS enzymes evolved prior to the divergence and global dispersal of the modern Asteraceae.

2. Material and methods

2.1. Identification and cloning of terpene synthases from *B. spinosa*

L. sativa germacrene A synthase 2 (*LsGAS2 or LTC2, AF489965*) was used as a BLAST query to identify putative *GAS* in *B. spinosa* using the available EST database on NCBI. The BLAST search rendered four putative *GAS* ESTs which cover the translated sequences sharing higher than 60% identity to that of LsGAS2. These ESTs included GE549025 (clone 1, 83% identity), GE549584 (clone 2, 71% identity), GE536704 (clone 3, 68% identity), and GE523822 (clone 4, 70% identity), and their corresponding clones were ordered from the Arizona Genomics Institute. After full-length sequencing of the ordered clones, clones 1, 2, and 3 were shown to constitute a single contiguous sequence that did not overlap with that of clone 4. These sequences were thus named *BsGAS1* (clones 1–3) and *BsGAS2* (clone 4). Since both 5′- and 3′-end sequences for *BsGAS1* and *BsGAS2* were not available, their full-length sequences were isolated using the SMART RACE

(Draft |
Accepted for publication in the Biochemical and Biophysical Research Communications
in September 2016)

cDNA Amplification Kit (Invitrogen) using the nested primers listed in Table S1 (primers 1-8; all primer sequences are given in Table S1 hereafter).

2.2. Identification of a third GAS isoform (LsGAS3) in *L. sativa*.

A new putative *germacrene A synthase* isoform from lettuce (*LsGAS3*) was identified using the sequence of BsGAS1 as query to search in the PhytoMetaSyn's Illumina-sequencing transcriptomic database [16,17]. This search resulted in a full-length cDNA sharing 80.5% identity with BsGAS1 upon translation. *LsGAS3* was amplified from lettuce cDNA using the primers 9/10. The amplicon was cloned into *pGEM T-easy* vector and sequenced.

2.3 Microbial Expression and kinetic analysis of recombinant BsGAS and LsGAS

2.3a *Production of sesquiterpenes in yeast*: *BsGAS1/2* genes were amplified using the primers 11/12 (*BsGAS1*) or primers 13/14 (*BsGAS2*). The corresponding amplicons were subsequently digested with *Bam*HI and *Xho*I, and inserted into their respective sites in *pESC-Leu2d* [18]. The resulted plasmids were transformed to the *Saccharomyces cerevisiae* EPY300 strain [19]. Sesquiterpene products were trapped in dodecane and analyzed by GC-MS as previously described [20].

2.3b *Expression of sesquiterpene synthases* in *Escherichia coli*: For bacterial expression, pDONR207 (Invitrogen) and pH9GW were used as the Gateway entry vector and destination vector, respectively. For protein expression, destination vector plasmids were transformed to BL21 (DE3). Cell cultures were grown at 37°C in Terrific Broth (TB) complemented with kanamycin (50mg/mL). Protein expression was induced, at OD600nm ≥ 0.8, with 0.1mM IPTG. The cultures were incubated with shaking for a further 5 h at 20°C. Cells were harvested by centrifugation. Cell pellets were re-suspended in 50 mL of buffer A (50 mM Tris–HCl, 50 mM glycine, 5% glycerol,

(Draft |
Accepted for publication in the Biochemical and Biophysical Research Communications
in September 2016)

0.5 M NaCl and 20 mM imidiazole, pH8), complemented with an EDTA-free protease inhibitor tablets (Sigma). The cells were lysed, and the clarified lysate was loaded onto a FPLC apparatus for a two-step purification using a 5mL Ni^{2+}-immobilized metal ion affinity chromatography column (HisTrap™ HP), equilibrated with buffer A, and an S200 26/60 Sephadex Gel filtration column equilibrated with 20 mM HEPES, 0.15 M NaCl, at pH7.5. Fractions containing purified protein were quantified using the Bradford assay, and protein purity was verified with SDS-PAGE.

2.3c *Kinetic characterization*: Kinetics were performed using the malachite green assay as previously described [21], using 96-well flat bottomed plates. The reaction was allowed in a total reaction volume of 50 µL consisting of malachite green assay buffer (25 mM MES, 25 mM CAPS, 50 mM Tris, 2.5 mU inorganic pyrophosphatase from *S. cerevisiae*, 5 mM $MgCl_2$ at pH 7), and serial dilutions of unlabeled farnesyl diphosphate FPP (100 to 1.6 µM). AaGAS, LsGAS2, BsGAS1, and BsGAS2 were added to final concentrations of 0.015, 0.025, 0.020 and 0.030 µM, respectively. Standard curves of monophosphate (Pi) and pyrophosphate (PPi) (50 µM to 0.01 µM) were generated using serial two-fold dilutions in malachite green assay buffer without FPP. Reactions were set up on ice in triplicate and incubated at 23 ^{o}C for 15 min. Enzyme reactions were quenched by addition of 12 µL of the malachite green development solution and incubated for 15 min prior readings at 623 nm on a Varioskan Flash plate-reader. The malachite green development solution was prepared as indicated previously [21]. In short, 300 mL of concentrated H_2SO_4 (18 M) was added to 1.5 L H_2O and equilibrated to room temperature. Malachite green powder (2.2 g) was then added and mixed with stirring. Prior to use, 10 mL of the malachite green dye was mixed with 2.5 mL of 7.5% ammonium molybdate and 0.2 mL 11% (v/v) Tween 20. Steady-state kinetic parameters (K_M, k_{cat} and k_{cat}/K_M) were obtained after non-linear regression analysis of the data, using the Michaelis-Menten model, with the GraphPad Prism.

(Draft |
Accepted for publication in the Biochemical and Biophysical Research Communications
in September 2016)

2.4 Phylogenetic analyses

A total of 28 terpene synthases including 23 GAS from across Asceraceae subfamilies (21) and non-Asteracea taxa (2), *Solidago candensis* germacrene D synthase, *Gossipium arboretum* δ-cadinene synthase, *Nicotiana tobacum* 5-*epi*-aristolochene synthase, and a monoterpene synthase were used for phylogenetic analysis (Table 2). The phylogeny of germacrene A oxidases (GAOs) was constructed using the sequences GAOs from *L. sativa* (LsGAO, GU198171), *Helianthus annuus* (HaGAO, GU256646), *Saussurea lappa* (SlGAO, GU256645), *B. spinosa* (BsGAO, GU256647), *Cichorium intybus* (CiGAO1, GU256644; and CiGAO2, HQ166835), and *Tanacetum parthenium* (TpGAO, KC964544). Premnaspirodiene oxygenase (HPO, EF569601) and 5-*epi*-aristolochene dihydroxylase from *Nicotiana tabacum* (EAH, AF368376) were also included in the GAO phylogenetic analysis. Sequences were aligned using the ClustalW algorithm with BLOSUM matrix. The phylogenetic trees was constructed based on the maximum likelihood method using the software MEGA 6.0 [22]. Bootstrap analysis was performed with 1,000 replicates to evaluate the statistical significance of each node.

3. Results and Discussion

3.1 Isolation of two *GAS* paralogues from *Barnadesia spinosa*

To investigate the presence of candidate GAS enzymes in *B. spinosa*, we used *LsGAS2* (Accession code: AF489965) sequence of *Lactuca sativa* (Cichorioideae subfamily) as a query in a BLAST search of available *B. spinosa* expressed sequence tag (EST) database on NCBI. Four EST clones encoding polypeptides showing 68-83% identity to LsGAS2 were identified. Sequencing and alignment of these EST clones revealed that three of them constitute a single contiguous sequence,

(Draft |
Accepted for publication in the Biochemical and Biophysical Research Communications
in September 2016)

suggesting that they are expressed from the same gene. The fourth EST clone showed homologous yet distinct sequence to the other three EST clones, indicating that it is transcribed from a different gene and that *B. spinosa* may encode two copies of putative GASs. Since both assembled sequences were partial, lacking both 5'- and 3'-ends of cDNAs, their full-length cDNAs were recovered by rapid-amplification of cDNA ends (RACE). Two *B. spinosa* transcripts named here as *BsGAS1* and *BsGAS2* were obtained (GenBank numbers KM066976 and KM066977, respectively; see functional characterization below). The isolated *BsGAS1* and *BsGAS2* cDNAs encode polypeptides of 580 and 554 amino acids, with predicted molecular masses of 66.5 and 64.1 kDa, respectively. BsGAS1 and BsGAS2 displayed 73.3% amino acid sequence identity to each other and showed 70.3% and 78.4% identity to LsGAS2 protein, respectively.

3.2. *In vivo* assessment of *B. spinosa* GAS enzymes

To examine the catalytic specificity of the enzymes encoded in *BsGAS1* and *BsGAS2*, open reading frames of the respective cDNAs were placed under the control of *Gal10* promoter and expressed in EPY300, a *Saccharomyces cerevisiae* yeast strain engineered to produce an elevated level of FPP endogenously [20]. As a control and standard, the previously characterized *LsGAS2* was expressed in parallel. C_{15}-isoprenoid volatiles from the culture were trapped with dodecane overlay and subsequently analyzed by gas chromatography-mass spectrometry (GC-MS). While the empty plasmid-transformed yeast only produced farnesol, an FPP product by an endogenous phosphatase activity, the yeast expressing *BsGAS1* or *BsGAS2* showed a single GC peak at ~16 min (Figure 2). Inspection of the EI mass spectra showed a parental ion with *m/z* value of 204 and main fragments including those with *m/z* values of 189, 161, 147, 93, and 68. Comparison of the fragmentation patterns revealed that the major compound from *BsGAS1*- or *BsGAS2*- expressing yeast were

(Draft |
Accepted for publication in the Biochemical and Biophysical Research Communications
in September 2016)

identical to each other, to published spectra from previously characterized GASs [12,23,24], and to the germacrene A standard from the *LsGAS2*-expressing yeast.

3.3. Kinetic characterization of BsGAS1 and BsGAS2

To assess the catalytic efficiency of *B. spinosa* GASs, recombinant BsGAS1 or BsGAS2 enzymes were produced and purified from *E. coli* as N-terminal histidine tag fusions. In parallel, lettuce GAS (LsGAS2) and *Artemisia annua* GAS (AaGAS) recombinant enzymes were prepared in an identical manner. The kinetic parameters of LsGAS2 and AaGAS enzymes have not been determined in previous studies. Known and fixed concentration of purified GAS enzymes were incubated with varying concentrations of FPP, and K_M and k_{cat} values were determined from Michaelis-Menten kinetic plots inferred from measuring the formation of pyrophosphates in the reaction solutions. Catalytic efficiency of LsGAS2 recombinant enzyme was the highest among four GAS enzymes (Table 1). Intriguingly, BsGAS1/2 and AaGAS enzymes showed comparable catalytic efficiencies to LsGAS2. These results indicate that *B. spinosa* and *A. annua* GAS retain potent GAS activity comparable to LsGAS2, although germacrene A-derived STLs have not been found from these two plant species.

3.4. Phylogenetic analysis of BsGAS1 and BsGAS2

To define the evolutionary relationship between the two paralogous *GASs* from *B. spinosa* and the broader collection of identified GAS enzymes, we constructed a phylogenetic tree (Figure 3A, Table 2). Inspection of the phylogenetic tree clearly indicates that GASs do not form a monophyletic group, suggesting convergent evolution for the germacrene A-synthesizing

(Draft |
Accepted for publication in the Biochemical and Biophysical Research Communications
in September 2016)

sesquiterpene synthases in the plant kingdom. However, most of the known GASs from the Asteraceae form a tight monophyletic group with the exception of GAS from Canadian goldenrod (*Solidago canadensis*). This result indicated that these GASs are descendants from a common ancestor in the Asteraceae. Of particular interest is that BsGAS1 and BsGAS2 fall into two distinct clades of GASs in the Asteraceae family as supported by strong bootstrap values (Figure 3A). BsGAS1 forms a cluster, named clade I, which includes GASs from chicory (*Cichorium intybus*) and *Ixeris dentana*, both of which belong to the Cichorieae tribe of the Cichorioideae subfamily. BsGAS2 forms another distinct group, named clade II, with other GASs from the Asteroideae, Cichorioideae, and Carduoideae subfamilies (Figure 3A). This result implies that clade-I GASs are retained in the Cichorieae tribe of the Cichorioideae subfamily but seem to disappear in other subfamilies, which constitutes the vast majority of the Asteraceae family.

In lettuce, both of the GAS paralogues previously characterized by Bennett *et al*. (2002) are in clade II. As lettuce is a member of the Cichorioideae subfamily, we hypothesized that there exists at least another clade-I GAS paralogue in this species. Using the clade-I BsGAS1 as query to search in the recently-released transcriptomic databases of medicinal plants [16], we identified a putative germacrene A synthase that shares 80.5% and 94.2% sequence identity with BsGAS1 and CiGASl, respectively. This new lettuce *GAS* paralogue was named *LsGAS3* (GenBank number: KU234689) and included in the phylogenetic analysis. Phylogenetic analysis showed that it was grouped with the clade-I GASs as predicted, further supporting our hypothesis that the Cichorioideae and the basal subfamily Barnadesioideae retain the clade-I GASs (Figure 3A).

Considering possible co-evolution of the next metabolic step in STL biosynthesis, we examined phylogenetic distribution of GAOs in relation to GASs (Figure 1B). Interestingly, our analysis revealed a similar pattern with two clusters similar to that of GASs (Figure 3B). One of

(Draft |
Accepted for publication in the Biochemical and Biophysical Research Communications
in September 2016)

the two clusters, named clade I, is conserved between the Cichorioideae and the Barnadesioideae but appear to be lost in other lineages – a situation that mirrors clade I of the GASs. These data suggest that both clade-I GASs and GAOs have been present in the Barnadesioideae subfamily since the last common ancestor of the Asteraceae, but disappeared in all except for the Cichorioideae subfamily. Retention of the clade-I GAS/GAO pair in the Cichorioideae may be associated with the evolution of certain specialized cell types, such as laticifers in the Cichorioideae, that are not present in other subfamilies. This intriguing link would be of great interest to investigate in follow-up studies.

The diversity of STLs, a signature family of specialized metabolites in the Asteraceae, likely contributed to the global distribution of this plant family, enabling an 'out of South America' migration once the biosynthetic potential for STLs was realized. Strikingly, the presence of a biochemically viable GAS in *B. spinosa* (as demonstrated here) indicates that the capacity to catalyze the first committed enzymatic step in STL biosynthesis evolved prior to subsequent adaptive evolution in the expanding Asteraceae lineage. Further, GAO activities are highly conserved throughout the Asteraceae, even at the evolutionary base; an observation that implies the presence of enzymatic machinery for the first oxidative transformation of germacrene A predates geographical expansion of the Asteraceae family. Given the clear congruencies of STL specialized metabolism and DNA sequence in the Asteraceae [25], the physical juxtaposition of GAS and GAO by genomic rearrangement was a likely scenario leading to synchronized expression and ultimately STL biosynthesis, consistent with the co-occurrence of TPS/P450 sets in sequenced plant genomes [26]. As a unit, the GAS/GAO tandem produces GAA leading to the simplest germacranolide (e.g. costunolide) – the non-enzymatic degradation and cyclization of which opens the gateway to diverse elemene, eudesmane and guaiene classes of STLs [5,27].

(Draft |
Accepted for publication in the Biochemical and Biophysical Research Communications
in September 2016)

Acknowledgements

This work was supported by National Science and Engineering Research Council of Canada (NSERC) to Ro, DK. This work was also supported by the joint Engineering and Physical Sciences Research Council/ Biotechnological and Biological Sciences Research Council (BBSRC)-funded OpenPlant Synthetic Biology Research Centre grant BB/L014130/1, the BBSRC Institute Strategic Programme Grant 'Understanding and Exploiting Plant and Microbial Metabolism' (BB/J004561/1) and the John Innes Foundation.

REFERENCES

[1] F.B. Da Costa, L. Terfloth, J. Gasteiger, Sesquiterpene lactone-based classification of three Asteraceae tribes: a study based on self-organizing neural networks applied to chemosystematics., Phytochemistry 66 (2005) 345-353.

[2] D. Hristozov, F.B. Costa, J. Gasteiger, Sesquiterpene lactone-based classification of the family Asteraceae using neural networks and k-nearest neighbours., J. Chem. Inf. Model 47 (2007) 9-19.

[3] V.D. Barreda, L. Palazzesi, L. Katinas, J.V. Crisci, M.C. Telleria, K. Bremer, M.G. Passalia, F. Bechis, R. Corsolini, An extinct Eocene taxon of the daisy family (Asteraceae): evolutionary, ecological and biogeographical implications, Ann. Bot. 109 (2012) 127-134.

[4] V.D. Barreda, L. Palazzesi, M.C. Telleria, L. Katinas, J.V. Crisci, K. Bremer, M.G. Passalia, R. Corsolini, R. Rodriguez Brizuela, F. Bechis, Eocene Patagonia fossils of the daisy family, Science 329 (2010) 1621.

[5] A.K. Picman, Biological activities of sesquiterpene lactone, Biochem. Syst. Ecol. 14 (1986) 255-281.

[6] K. Bremer, Subfamily Barnadesioideae, in: K. Bremer (Ed.), Astaraceae. Cladistics and Classification, Timber Press, Portland, 1994, pp. 49-60.

[7] M. Gruenstaeudl, E. Urtubey, R.K. Jansen, R. Samuel, M.H. Barfuss, T.F. Stuessy, Phylogeny of Barnadesioideae (Asteraceae) inferred from DNA sequence data and morphology, Mol. Phylogenet. Evol. 51 (2009) 572-587.

[8] K. Bremer, R.K. Jansen, A new subfamily of the Asteraceae, Ann. Missouri Bot. Garden 79 (1992) 414-415.

[9] R.K. Jansen, J.D. Palmer, A chloroplast DNA inversion marks an ancient evolutionary split in the sunflower family (Asteraceae), Proc. Natl. Acad. Sci. USA 84 (1987) 5818-5822.

[10] B.A. Bohm, T.F. Stuessy, Flavonoids of Mutisieae and Barnadesioideae. In Flavonoids of the Sunflower Family (Asteraceae), Springer-Verlag, 2001, pp. 340-346.

[11] F.C. Seaman, Sesquiterpene lactones as taxonomic characters in the Asteraceae, Bot. Rev. 48 (1982) 121-594.

[12] M.H. Bennett, J.W. Mansfield, M.J. Lewis, M.H. Beale, Cloning and expression of sesquiterpene synthase genes from lettuce (*Lactuca sativa* L.), Phytochemistry 60 (2002) 255-261.

[13] N. Ikezawa, J.C. Gopfert, D.T. Nguyen, S.U. Kim, P.E. O'Maille, O. Spring, D.K. Ro, Lettuce costunolide synthase (CYP71BL2) and its homolog (CYP71BL1) from sunflower catalyze distinct regio- and stereoselective hydroxylations in sesquiterpene lactone metabolism, J. Biol. Chem. 286 (2011) 21601-21611.

[14] Q. Liu, M. Majdi, K. Cankar, M. Goedbloed, T. Charnikhova, F.W. Verstappen, R.C. de Vos, J. Beekwilder, S. van der Krol, H.J. Bouwmeester, Reconstitution of the costunolide biosynthetic pathway in yeast and *Nicotiana benthamiana*, PLoS One 6 (2011) e23255.

[15] D.T. Nguyen, J.C. Gopfert, N. Ikezawa, G. Macnevin, M. Kathiresan, J. Conrad, O. Spring, D.K. Ro, Biochemical conservation and evolution of germacrene A oxidase in Asteraceae, J. Biol. Chem. 285 (2010) 16588-16598.

[16] P.J. Facchini, J. Bohlmann, P.S. Covello, V.D. Luca, R. Mahadevan, J.E. Page, D.K. Ro, C.W. Sensen, R. Storms, V.J.J. Martin, Synthetic biosystems for the production of high-value plant metabolites, Trends Biotechnol. 30 (2012) 127-131.

[17] M. Xiao, Y. Zhang, X. Chen, E.-J. Lee, C.J.S. Barber, R. Chakrabarty, I. Desgagné-Penix, T.M. Haslam, Y.-B. Kim, E. Liu, G. MacNevin, S. Masada-Atsumi, D.W. Reed, J.M. Stout, P. Zerbe, Y. Zhang, J. Bohlmann, P.S. Covello, V.D. Luca, J.E. Page, D.K. Ro, V.J.J. Martin, P.J. Facchini, C.W. Sensen, Transcriptome analysis based on next-generation sequencing of non-model plants producing specialized metabolites of biotechnological interest, J. Biotechnol. 166 (2013) 122-134.

[18] D.K. Ro, M. Ouellet, E.M. Paradise, H. Burd, D. Eng, C.J. Paddon, J.D. Newman, J.D. Keasling, Induction of multiple pleiotropic drug resistance genes in yeast engineered to produce an increased level of anti-malarial drug precursor, artemisinic acid, BMC Biotechnol. 8 (2008) 83.

[19] D.K. Ro, E.M. Paradise, M. Ouellet, K.J. Fisher, K.L. Newman, J.M. Ndungu, K.A. Ho, R.A. Eachus, T.S. Ham, J. Kirby, M.C. Chang, S.T. Withers, Y. Shiba, R. Sarpong, J.D. Keasling, Production of the antimalarial drug precursor artemisinic acid in engineered yeast, Nature 440 (2006) 940-943.

[20] T.D. Nguyen, G. MacNevin, D.K. Ro, De novo synthesis of high-value plant sesquiterpenoids in yeast, Methods Enzymol. 517 (2012) 261-278.

[21] M. Vardakou, M. Salmon, J.A. Faraldos, P. E. O'Maille, Comparative analysis and validation of the malachite green assay for the high throughput biochemical characterization of terpene synthases, MethodsX (2014) 187-196.

[22] K. Tamura, G. Stecher, D. Peterson, A. Filipski, S. Kumar, MEGA6: Molecular Evolutionary Genetics Analysis version 6.0, Mol. Biol. Evol. 30 (2013) 2725-2729.

[23] H.J. Bouwmeester, J. Kodde, F.W. Verstappen, I.G. Altug, J.W. de Kraker, T.E. Wallaart, Isolation and characterization of two germacrene A synthase cDNA clones from chicory, Plant Physiol. 129 (2002) 134-144.

[24] L. Pazouki, H.R. Memari, A. Kannaste, R. Bichele, U. Niinemets, Germacrene A synthase in yarrow (*Achillea millefolium*) is an enzyme with mixed substrate specificity: gene cloning, functional characterization and expression analysis, Front. Plant Sci. 6 (2015) 111.

[25] C. Ekenäs, J. Rosén, S. Wagner, I. Merfort, A. Backlund, K. Andreasen, Secondary chemistry and ribosomal DNA data congruencies in Arnica(Asteraceae), Cladistics 25 (2009) 78-92.

[26] A.M. Boutanaev, T. Moses, J. Zi, D.R. Nelson, S.T. Mugford, R.J. Peters, A. Osbourn, Investigation of terpene diversification across multiple sequenced plant genomes, Proc. Natl. Acad. Sci. USA 112 (2015) e81-88.

[27] G.F. Padilla-Gonzalez, F.A. dos Santos, F.B. Da Costa, Sesquiterpene lactones: more than protective plant compounds with high toxicity, Crit. Rev. Plant Sci. 35 (2016) 18-37.

Figure Legends

Figure 1. Taxonomic and chemical diversity of the Asteraceae.

A. A simplified phylogeny of the Asteraceae family based on the study of Panero and Funk (2008). Asterisks indicate the basal lineage (Barnadesioideae) and three major subfamilies. B. The biosynthetic pathway of germacrene A-based sesquiterpene lactones in the Asteraceae. FPP: farnesyl pyrophosphate. Solid and dashed arrows represent biochemically characterized and unknown steps, respectively.

Figure 2. GC–MS analysis of the metabolites produced by EPY300 yeast expressing *B. spinosa* and lettuce *GASs.*

Figure 3. Phylogenies of germacrene A synthases (A) and germacrene A oxidases (B) in the Asteraceae family.

The evolution history was inferred based on the maximum likelihood method and the JTT matrix-based model. The percentage of trees in which the associated taxa clustered together is shown next to the branches. The analysis involved 26 protein sequences (Table 2). All positions containing gaps and missing data were eliminated. The branch lengths are measured in the number of substitutions per site. Brackets indicate clades of GASs and GAOs. Enzymes from the Barnadesioideae and the Cichorioideae are indicated by bold and underlined labels, respectively. Non-Asteraceae enzymes are framed.

(Draft |
Accepted for publication in the Biochemical and Biophysical Research Communications
in September 2016)

Table 1. Kinetic properties of four recombinant GASs from the Asteraceae.

	k_{cat} (s^{-1})	K_M (μM)	k_{cat}/K_M ($s^{-1}\mu M^{-1}$)
[a]LsGAS2	0.28 ± 0.03	10.9 ± 0.16	0.026
[a]BsGAS1	0.13 ± 0.01	7.8 ± 0.25	0.017
[a]BsGAS2	0.28 ± 0.02	14.8 ± 0.29	0.019
[a]AaGAS	0.25 ± 0.02	10.8 ± 0.32	0.023

[a]Ls: *Lactuca sativa*, Bs: *Barnadesia spinosa*, Aa: *Artemisia annua*

Table 2. Germacrene A synthase and other sesquiterpene synthase sequences used in the phylogenetic analysis. Non-Asteraceae sources are indicated in brackets next to species names.

Species	GenBank number	Abbreviation
Barnadesia spinosa	KM066976	BsGAS1
Barnadesia spinosa	KM066977	BsGAS2
Artemisia annua	DQ447636	AaGAS
Lactuca sativa	AF489964	LsGAS1
Lactuca sativa	AF489965	LsGAS2
Lactuca sativa	KU234689	LsGAS3
Cichorium intybus	AF498000	CiGAS.s (short)
Cichorium intybus	AF497999	CiGAS.l (long)
Matricaria chamomilla	JQ255377	McGAS
Helianthus annuus	DQ016667	HaGAS1
Helianthus annuus	EU327785	HaGAS2
Helianthus annuus	GU176380	HaGAS3
Ixeris dentata	AY082672	IdGAS
Solidago canadensis	AJ304452	ScGAS
Solidago canadensis	AJ583447	ScGDS
Crepidiastrum sonchifolium	DQ186657	CsGAS
Cynara cardunculus	JN383985	CcGAS
Tanacetum cinerariifolium	KC441526	TcGAS
Tanacetum parthenium	JF819848	TpGAS
Achillea millefolium	KC145534	AmGAS
Pogostemon cablin (Lamiaceae)	AY508728	PcGAS
Vitis vinifera (Vitaceae)	HQ326230	VvGAS
Nicotiana tabaccum (Solanaceae)	L04680	NtEAS
Gossypium arboretum (Malvaceae)	U23206	GaCDS
Abies grandis (Pinaceae)	U87909	AgPS

Note: GAS: germacrene A synthase. GDS: germacrene D synthase, CDS: cadinene synthase, EAS: 5-*epi*-aristolochene synthase, PS: pinene synthase.

(Draft |
Accepted for publication in the Biochemical and Biophysical Research Communications
in September 2016)

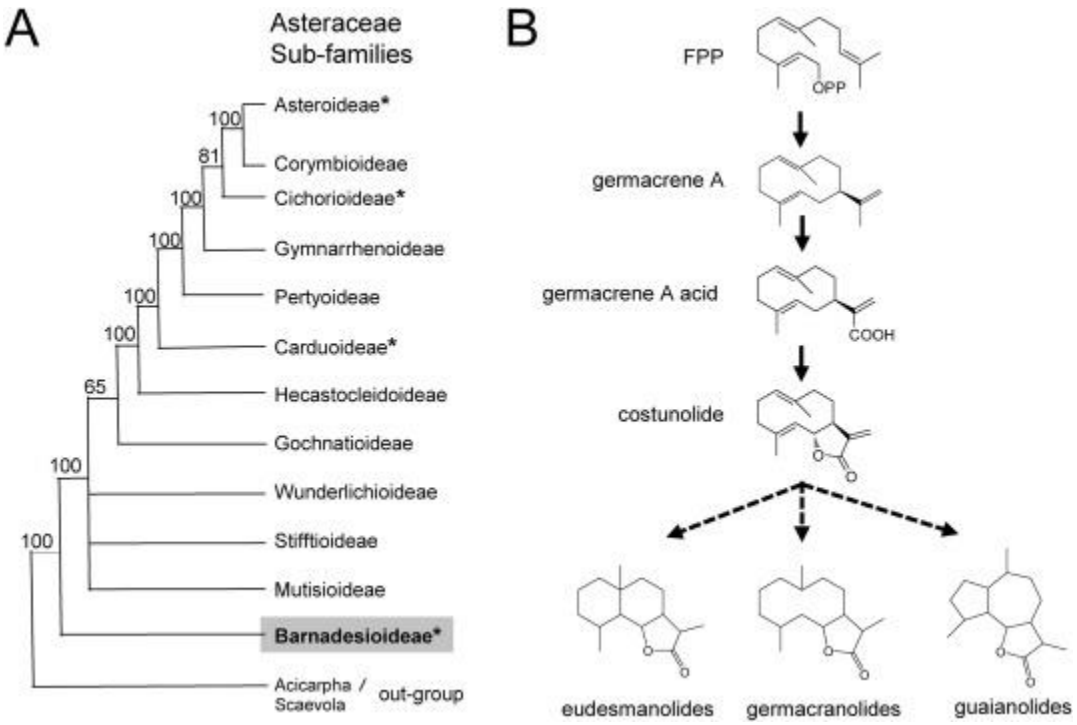

Figure 1.

(Draft |
Accepted for publication in the Biochemical and Biophysical Research Communications
in September 2016)

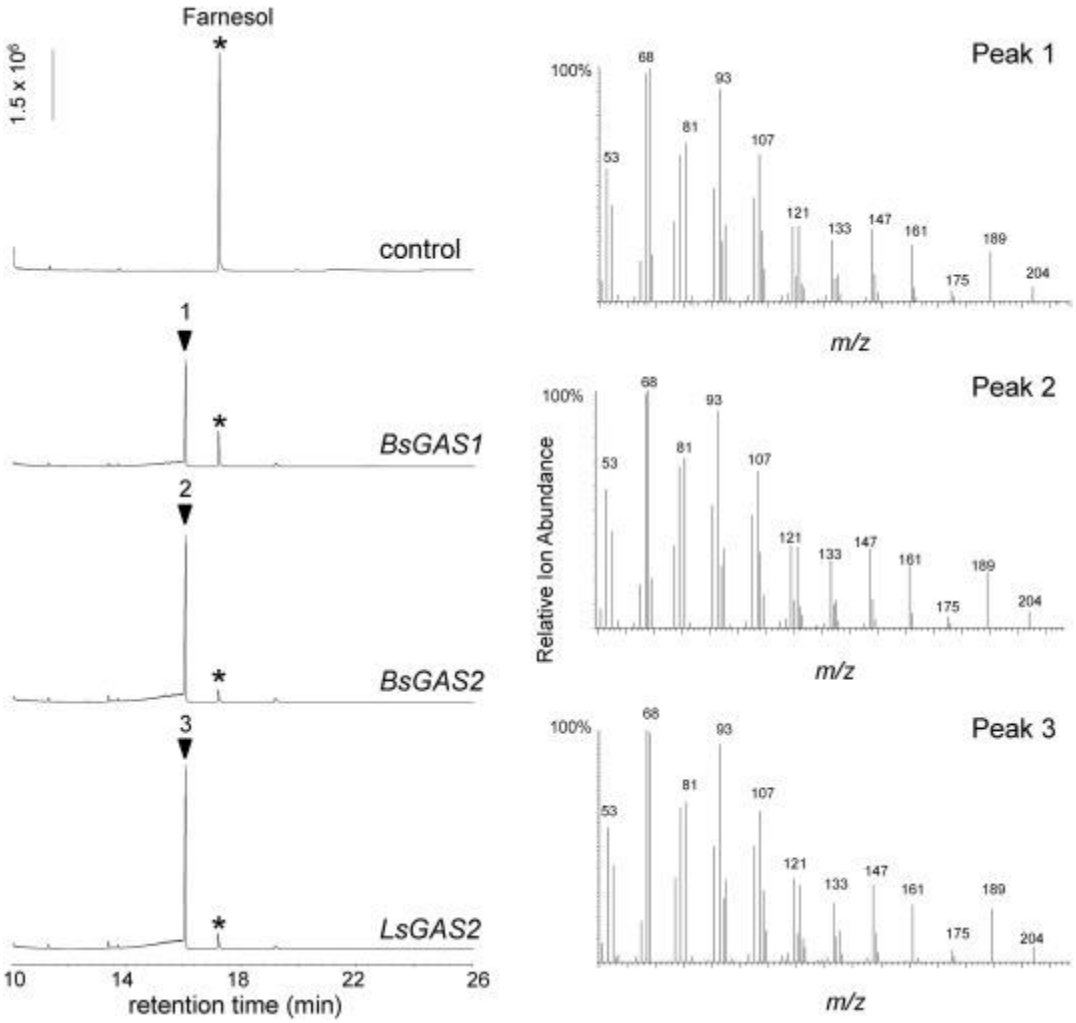

Figure 2.

(Draft |
Accepted for publication in the Biochemical and Biophysical Research Communications
in September 2016)

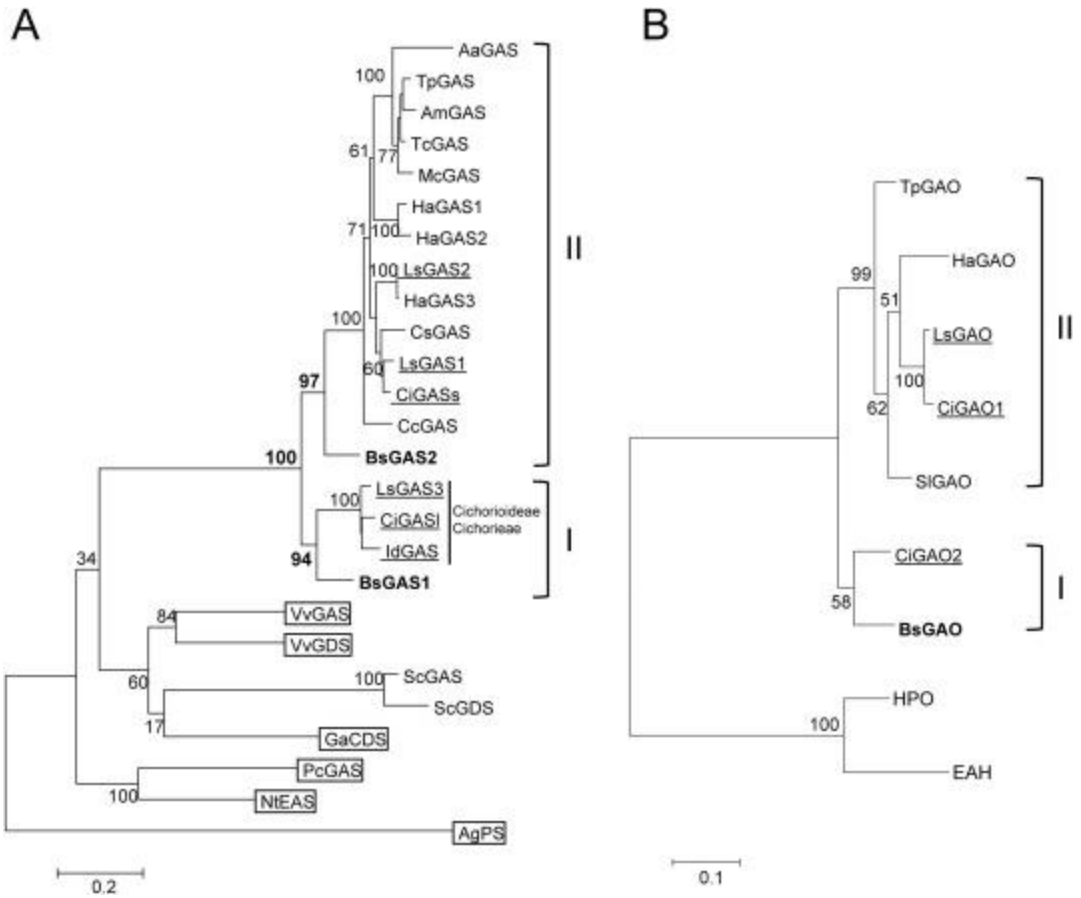

Figure 3.

(Draft |
Accepted for publication in the Biochemical and Biophysical Research Communications
in September 2016)