Title: Did a plant-herbivore arms race drive chemical diversity in *Euphorbia?*

Authors: M. Ernst[1,2,3], L.-F. Nothias[2,3], J. J. J. van der Hooft[2,3,4], R. R. Silva[2,3], C. H. Saslis-Lagoudakis[1], O. M. Grace[5], K. Martinez-Swatson[1], G. Hassemer[1], L. A. Funez[7], H. T. Simonsen[6], M. H. Medema[4], D. Staerk[8], N. Nilsson[9], P. Lovato[9], P. C. Dorrestein[2,3,10*] & N. Rønsted[1*]

Affiliations:

[1]Natural History Museum of Denmark, Faculty of Science, University of Copenhagen, Øster Farimagsgade 5, DK-1353 Copenhagen K, Denmark.

[2]Collaborative Mass Spectrometry Innovation Center, Skaggs School of Pharmacy and Pharmaceutical Sciences, University of California, San Diego, La Jolla, CA 92093, USA.

[3]Skaggs School of Pharmacy and Pharmaceutical Sciences, University of California, San Diego, La Jolla, CA 92093, USA.

[4]Bioinformatics Group, Department of Plant Sciences, Wageningen University, 6708 PB, Wageningen, The Netherlands.

[5]Comparative Plant & Fungal Biology, Royal Botanic Gardens, Kew, Surrey TW9 3AE, United Kingdom.

[6]Department of Biotechnology and Biomedicine, Technical University of Denmark, Søltofts Plads, DK-2800 Kgs. Lyngby, Denmark.

[7]Herbário Dr. Roberto Miguel Klein (FURB), Regional University of Blumenau, Campus 1, 89012-900, Blumenau, SC, Brazil.

[8]Department of Drug Design and Pharmacology, University of Copenhagen, Universitetsparken 2, DK-2100 Copenhagen, Denmark.

[9]Front End Innovation, LEO Pharma A/S, Industriparken 55, DK-2750 Ballerup, Denmark.

[10]Center for Microbiome Innovation, University of California, San Diego, La Jolla, CA 92093, USA.

*Correspondence to: nronsted@snm.ku.dk and pdorrestein@ucsd.edu

Abstract:

The genus *Euphorbia* is among the most diverse and species-rich plant genera on Earth, exhibiting a near-cosmopolitan distribution and extraordinary chemical diversity, especially across highly toxic macro- and polycyclic diterpenoids. However, very little is known about drivers and evolutionary origins of chemical diversity within *Euphorbia*. Here, we investigate 43 *Euphorbia* species to understand how geographic separation over evolutionary time has impacted chemical differentiation. We show that the structurally highly diverse *Euphorbia* diterpenoids are significantly reduced in species native to the Americas, compared to the Eurasian and African continents, where the genus originated. The localization of these compounds to young stems and roots suggest ecological relevance in herbivory defense and immunomodulatory defense mechanisms match diterpenoid levels, indicating chemo-evolutionary adaptation to reduced herbivory pressure.

One Sentence Summary:

Global chemo-evolutionary adaptation of *Euphorbia* affected immunomodulatory defense mechanisms.

Main Text:

Euphorbia is among the most diverse and species-rich plant genera on Earth, exhibiting a near-cosmopolitan distribution and extraordinary chemical diversity among 2,000 species (*1-3*). The genus originated in Africa approximately 48 million years ago and through two single long-distance dispersal events 30 and 25 million years ago, expanded to the American continents (Fig. 1) (*1-2*). *Euphorbia* chemical diversity is characterized by an extraordinary diversity of macro- and polycyclic diterpenoids, biosynthetically derived from a head-to-tail cyclization of the tetraprenyl pyrophosphate precursor (*3, 4*). These compounds play an important ecological role as feeding deterrents and have shown exclusive occurrence and chemotaxonomic relevance in the plant families Euphorbiaceae and Thymelaceae (*3, 5-8*). However, the chemo-evolutionary transitions driving chemical diversity within *Euphorbia* are unknown. Here, we investigate the role of biogeography in the evolution of specialized metabolite diversity in 43 *Euphorbia* species, representing the genus' global genetic diversity and biogeographic history across all continents.

Coevolutionary theory suggests that an arms race between plants and herbivores yields increased specialized metabolite diversity (*9-11*). The evolution of a chemically different and biologically more active molecule increases a plant's fitness by reducing the fitness of its predator (*10, 12*), and the probability of producing one or more biologically active compounds may increase with phytochemical diversity (*13*). To assess specialized metabolite diversity in relation to the evolutionary and biogeographic history of *Euphorbia*, we subjected extracts of 43 *Euphorbia* species to liquid chromatography tandem mass spectrometry (LC-MS/MS), created mass spectral molecular networks through Global Natural Products Social Molecular Networking (GNPS) (*14, 15*) and calculated the chemical structural and compositional similarity (CSCS) for all *Euphorbia* subgeneric clades (*16*). Our data show significantly higher chemical similarity among species of subgenus *Chamaesyce* compared to the mean chemical similarity among species of the remaining subgeneric clades (Fig. 2). The only species

clustered in the chemogram (sharing high chemical similarity) are 8 out of 9 species representing the American clade within subgenus *Chamaesyce* (Fig. 2D). Consistent with the coevolutionary theory, the reduction of chemical structural diversity in these species suggests an adaptation to reduced herbivory pressure in the Americas during the biogeographic history of the genus.

To further understand the chemo-evolutionary relationships of *Euphorbia* at a molecular level we putatively identified major specialized metabolite classes by combining mass spectral molecular networking with *in silico* annotation tools (*17-19*), substructure recognition (*20, 21*) and automated chemical classification through ClassyFire (*22*). This resulted in annotated compound classes for over 30% of the compounds detected (23, *24*) (Fig. S1 and S2, supplementary text). Our approach revealed many known *Euphorbia* diterpenoids as well as other metabolite classes. Out of six major structural classes of *Euphorbia* specialized metabolites (sesquiterpenoids, diterpenoids, cerebrosides, phenolics, flavonoids, and triterpenoids including steroids), four are found in our molecular networks (i.e., diterpenoids; triterpenoids including cholestane and ergostane steroids; steroid lactones and stigmastanes; and glycosylglycerols corresponding to cerebrosides). Additionally, *in silico* structure annotation suggests the presence of tricarboxylic and benzoic acids and derivatives as well as fatty alcohols and glycosphingolipids (Fig. 2A). Among the *Euphorbia* diterpenoids, we observe different skeletal types within the same molecular families (two or more connected components of a graph) (Fig. 3). Many *Euphorbia* diterpenoid backbone skeletons are isomeric, and their respective fragmentation spectra are highly similar (*25*). Nonetheless, we are able to distinguish different diterpene spectral fingerprints within a molecular family by mapping Mass2Motifs on the mass spectral molecular networks. Mass2Motifs correspond to common patterns of mass fragments and neutral losses, which are extracted using unsupervised substructure discovery of the MS/MS data through MS2LDA (*20, 21*). Combining *in silico* structure annotation, automated chemical classification, and MS2LDA, allows us to putatively identify chemical classes within the mass spectral molecular networks, as well as chemical subclasses within molecular families (Fig. 3, Fig. S2). Consistent with previous observations of *Euphorbia* diterpenoids exhibiting anti-herbivore biological activity (*5-8*) and the low chemical diversity exhibited by subgenus *Chamaesyce*, we find very few to no *Euphorbia* diterpenoids in representatives of the American radiation of subgenus *Chamaesyce* (Fig. 1). Although subgenus *Chamaesyce* includes the largest American radiation within the genus *Euphorbia*, there is also a smaller American radiation within subgenus *Euphorbia* (Fig. 1). The two investigated representatives of this clade contain intermediate amounts of *Euphorbia* diterpenoids. The clade is estimated to have originated approximately 5 million years later than the American radiation within subgenus *Chamaesyce* (*1*), which could suggest less time for adaptation to reduced herbivory pressure.

To understand where the *Euphorbia* diterpenoids are produced within the plants, we dissected four species representing the four subgeneric clades into approximately 20 sections (*23*) (Fig. S10, Fig. S11). Mass spectrometric investigation revealed that diterpenoids are primarily found in the roots, in representatives of subgenera *Euphorbia* and *Athymalus* (*E. milii* var. *hislopii* and *E. horrida,* Fig. 4, Fig. S3, Fig. S6-S9). In the European subgenus *Esula* (*E. lathyris*), diterpenoid production is also pronounced in other plant parts, such as the young

stems (Fig. 4, Fig. S5). Consistent with the lower chemical diversity reported above, diterpenoid production is reduced or absent from most sections throughout the whole plant in *E. hirta*, a representative of the American clade within subgenus *Chamaesyce* (Fig 4, Fig. S4, Fig. S10). Compartmentalization of the diterpenoids to mainly young stems and roots underpins their function as anti-feeding molecules (*5*-*8*).

As anti-herbivore activity cannot be directly tested from the past continental transition, we set out to test a bioactivity that could reflect a chemical defense mechanism. One strategy of defense invoked by plants to overcome their sessile habit, is through immunomodulatory effects on herbivores. *Euphorbia* diterpenoids are known to exhibit immunomodulatory activities through the selective modulation of protein kinase C (PKC) (*26, 27*). Therefore, we evaluated the modulation of PKC by measuring the capacity of small extract fractions, corresponding to compounds or compound groups in each of the 43 *Euphorbia* species, that modulate *in vitro* TNF-α release from peripheral blood mononuclear cells (PBMCs). To pinpoint the role of *Euphorbia* diterpenoids involved with TNF-α modulating properties, we tested for correlation between the number of bioactive fractions and the number of molecules within the previously annotated chemical classes using phylogenetic generalized least squares regression analysis (PGLS). The association between the number of TNF-α modulating fractions and the number of *Euphorbia* diterpenoids is significant (*P*-value: 0.02) (Fig. 1B, Fig. 1C, Table S1). Besides the *Euphorbia* diterpenoids, we also observe significant associations (*P*-value < 0.05) between the number of TNF-α modulating fractions and the number of overall diterpenoids and glycosyl glycerols with the best fit observed for the *Euphorbia* diterpenoids (Fig. 1B, Fig. 1C, Fig. S12, Table S1), supporting our hypothesis of the ecological function of *Euphorbia* diterpenoids as immunomodulatory defense molecules.

To evaluate the possibility of *Euphorbia* diterpenoids being produced as a response to local plant-predator interactions, we compiled a dataset of known *Euphorbia* herbivores from the literature (*23*). Several hawkmoth species of the genus *Hyles* were found to be highly specialized predators of *Euphorbia* (*5, 28*). They have been shown to exhibit host specificity and to tolerate the highly toxic *Euphorbia* diterpenoids, which they reuse as a defense strategy against their own predators by regurgitating plant material from the gut (*5, 8*). Native species distribution data suggests a close co-occurrence of *Euphorbia*-feeding *Hyles* species with the chemically highly diverse and biologically active European species of subgenus *Esula* (Fig. 1A), and an absence in the American habitats of the chemically less diverse and biologically little active representatives of subgenus *Chamaesyce*. However, our data also suggests that African members of subgenus *Athymalus* and subgenus *Euphorbia* occurring in Southern Africa and Madagascar, outside of the distribution range of *Euphorbia*-feeding *Hyles*, produce a high diversity of feeding deterrent diterpenoids (Fig. 1). Thus, we speculate, that previously not described (or extinct) generalist or specialist *Euphorbia*-feeding herbivores occur (or occurred) in these regions, which contributed to maintaining adaptive pressure. The black rhinoceros, *Diceros bicornis* L. distributed in the southern African subregion, for example, was found to feed often and extensively on African *Euphorbia* species of subgenus *Euphorbia* (*29*). Anecdotal evidence of the lack of specialized herbivores in the Americas supports our results. A single hawkmoth species, *Hyles euphorbiae*, was only introduced recently to North America (*5*), where it was used as a host-specific enemy and biological control for the

reportedly highly invasive European species of subgenus *Esula* (*E. cyparissias* and *E. esula*), lacking predators in the newly occupied habitats. Furthermore, although the poinsettia (*E. pulcherrima* Willd. ex Klotzsch), a very well-known house plant and American representative of subgenus *Chamaesyce,* is notorious for its "extreme toxicity" among the general public, toxicity remains unconfirmed in the clinic, as 92.4% of patients exposed to the plant did not develop adverse effects (*30*), corroborating the low chemical diversity and immunomodulatory activity observed here.

In contempt of the limited knowledge about *Euphorbia* herbivory, the remarkable immunomodulatory activities of the *Euphorbia* diterpenoids provide an indirect way to assess chemical defense against predators. The differential biosynthesis of diterpenoids in species geographically separated through evolutionary time, suggests differential exposure to herbivory during the biogeographic history of the genus. Indeed, there are no known herbivores of the American species, while specialized herbivores are well documented for the European and African species (*8, 29*). The mechanism of predator tolerance is not known, but the presence of specialized herbivores is consistent with our results and hypothesis that the greater the diversity of herbivores feeding on a plant species, the more immunomodulating molecules the plants produce.

References and Notes:

1. J. W. Horn, Z. Xi, R. Riina, J. A. Peirson, Y. Yang, B. L. Dorsey, P. E. Berry, K. J. Wurdack, Evolutionary bursts in *Euphorbia* (Euphorbiaceae) are linked with photosynthetic pathway. *Evolution* **68**, 3485–3504 (2014).
2. J. W. Horn, B. W. van Ee, J. J. Morawetz, R. Riina, V. W. Steinmann, P. E. Berry, K. J. Wurdack, Phylogenetics and the evolution of major structural characters in the giant genus *Euphorbia* L. (Euphorbiaceae). *Mol. Phylogenetics Evol.* **63**, 305–326 (2012).
3. A. Vasas, J. Hohmann, *Euphorbia* diterpenes: isolation, structure, biological activity, and synthesis (2008-2012). *Chem. Rev.* **114**, 8579-8612 (2014).
4. Q. W. Shi, X. H. Su, H. Kiyota, Chemical and pharmacological research of the plants in genus *Euphorbia*. *Chem. Rev.* **108**, 4295-4327 (2008).
5. S. W. T. Batra, Establishment of *Hyles euphorbiae* (L.) (Lepidoptera: Sphingidae) in the United States for control of the weedy spurges *Euphorbia esula* L. and *E. cyparissias* L. *J. N. Y. Entomol. Soc.* **91**, 304–311 (1983).
6. J. Hua, Y. Liu, C.-J. Xiao, S.-X. Jing, S.-H. Luo, S.-H. Li, Chemical profile and defensive function of the latex of *Euphorbia peplus*. *Phytochemistry* **136**, 56-64 (2017).
7. C.-S. Huang, S.-H. Luo, Y.-L. Li, C.-H. Li, J. Hua, Y. Liu, S.-X. Jing, Y. Wang, M.-J. Yang, S.-H. Li, Antifeedant and antiviral diterpenoids from the fresh roots of *Euphorbia jolkinii*. *Nat. Prod. Bioprospect.* **4**, 91-100 (2014).
8. A. K. Hundsdoerfer, J. N. Tshibangu, B. Wetterauer, M. Wink, Sequestration of phorbol esters by aposematic larvae of *Hyles euphorbiae* (Lepidoptera: Sphingidae)? *Chemoecology* **15**, 261–267 (2005).
9. J. X. Becerra, The impact of herbivore–plant coevolution on plant community structure. *Proc. Natl. Acad. Sci. U S A* **104**, 7483–7488 (2007).

10. P. R. Ehrlich, P. H. Raven, Butterflies and plants: A study in coevolution. *Evolution* **18**, 586–608 (1964).
11. L. A. Richards, L. A. Dyer, M. L. Forister, A. M. Smilanich, C. D. Dodson, M. D. Leonard, C. S. Jeffrey. Phytochemical diversity drives plant–insect community diversity. *Proc. Natl. Acad. Sci. U S A* **112**, 10973–10978 (2015).
12. R. D. Firn, C. G. Jones, Natural products - a simple model to explain chemical diversity. *Nat. Prod. Rep.* **20**, 382–391 (2003).
13. C. G. Jones, R. D. Firn, S. B. Malcolm, On the evolution of plant secondary chemical diversity. *Phil. Trans. R. Soc. Lond. B.* **333**, 273-280 (1991).
14. M. Wang et al., Sharing and community curation of mass spectrometry data with Global Natural Products Social Molecular Networking. *Nat. Biotechnol.* **34**, 828-837 (2016).
15. J. Watrous, P. Roach, T. Alexandrov, B. S. Heath, J. Y. Yang, R. D. Kersten, M. van der Voort, K. Pogliano, H. Gross, J. M. Raaijmakers, B. S. Moore, J. Laskin, N. Bandeira, P. C. Dorrestein, Mass spectral molecular networking of living microbial colonies. *Proc. Natl. Acad. Sci. U S A* **109**, E1743–E1752 (2012).
16. B. E. Sedio, J. C. Rojas Echeverri, P. Boya, S. J. Wright, Sources of variation in foliar secondary chemistry in a tropical forest tree community. *Ecology* **98**, 616–623 (2017).
17. K. Dührkop, S. Böcker, Fragmentation trees reloaded. *J. Cheminform.* **8**, 5 (2016).
18. K. Dührkop, H. Shen, M. Meusel, J. Rousu, S. Böcker, Searching structure databases using CSI:FingerID. *Proc. Natl. Acad. Sci. U S A* **112**, 12580-12585 (2015).
19. R. R. da Silva, M. Wang, L.-F. Nothias, J. J. van der Hooft, A. M. Caraballo- Rodriguez, E. Fox, M. J. Balunas, J. L. Klassen, N. P. Lopes, P. C. Dorrestein, Propagating annotations of molecular networks using in silico fragmentation. *PLos Comput. Biol.* **14**, e1006089 (2018).
20. J. J. J. van der Hooft, J. Wandy, M. P. Barrett, K. E. V. Burgess, S. Rogers, Topic modeling for untargeted substructure exploration in metabolomics. *Proc. Natl. Acad. Sci. U S A* **113**, 13738-13743 (2016).
21. J. Wandy, Y. Zhu, J. J. J. van der Hooft, R. Daly, M. P. Barrett, S. Rogers, Ms2lda.org: web-based topic modelling for substructure discovery in mass spectrometry. *Bioinformatics* **34**, 317-318 (2018).
22. Y. Djoumbou Feunang, R. Eisner, C. Knox, L. Chepelev, J. Hastings, G. Owen, E. Fahy, C. Steinbeck, S. Subramanian, E. Bolton, R. Greiner, D. S. Wishart, Classyfire: automated chemical classification with a comprehensive, computable taxonomy. *J. Cheminform.* **8**, 61 (2016).
23. Materials and methods are available as supplementary materials.
24. L. W. Sumner, A. Amberg, D. Barrett, M. H. Beale, R. Beger, C. A. Daykin, T. W.-M. Fan, O. Fiehn, R. Goodacre, J. L. Griffin, T. Hankemeier, N. Hardy, J. Harnly, R. Higashi, J. Kopka, A. N. Lane, J. C. Lindon, P. Marriott, A. W. Nicholls, M. D. Reily, J. J. Thaden, M. R. Viant, Proposed minimum reporting standards for chemical analysis Chemical Analysis Working Group (CAWG) Metabolomics Standards Initiative (MSI). *Metabolomics* **3**, 211-221 (2007).
25. L.-F. Nothias-Scaglia, I. Schmitz-Afonso, F. Renucci, F. Roussi, D. Touboul, J. Costa, M. Litaudon, J. Paolini, Insights on profiling of phorbol, deoxyphorbol, ingenol and jatrophane diterpene esters by high performance liquid chromatography coupled to multiple stage mass spectrometry. *J. Chromatogr. A* **1422**, 128–139 (2015).

26. B. Berman, New developments in the treatment of actinic keratosis: focus on ingenol mebutate gel. *Clin. Cosmet. Investig. Dermatol.* **20**, 111–122 (2012).
27. S. E. Smith, B. S. Meldrum, The protein kinase C activators, phorbol 12-myristate,13-acetate and phorbol 12,13-dibutyrate, are convulsant in the pico- nanomolar range in mice. *Eur. J. Pharmacol.* **213**, 133-135 (1992).
28. A. K. Hundsdoerfer, D. Rubinoff, M. Attié, M. Wink, I. J. Kitching, A revised molecular phylogeny of the globally distributed hawkmoth genus *Hyles* (Lepidoptera: Sphingidae), based on mitochondrial and nuclear DNA sequences. *Molecular Phylogenetics and Evolution* **52**, 852-865 (2009).
29. B. D. Loutit, G. N. Louw, M. K. Seeley, First approximation of food preferences and the chemical composition of the diet of the desert-dwelling black rhinoceros, *Diceros bicornis* L. *Madoqua* **15**, 35-54 (1987).
30. E. P. Krenzelok, T. D. Jacobsen, J. M. Aronis, Poinsettia exposures have good outcomes… just as we thought. *Am. J. Emerg. Med.* **14**, 671-674 (1996).
31. G. Vogg, S. Achatz, A. Kettrup, H. Sandermann Jr., Fast, sensitive and selective liquid chromatographic-tandem mass spectrometric determination of tumor-promoting diterpene esters. *J. Chromatogr. A* **855**, 563-573 (1999).
32. G. Appendino, "Ingenane diterpenoids" in *Progress in the Chemistry of Organic Natural Products 102*, A. D. Kinghorn, H. Falk, S. Gibbons, J. Kobayashi, Eds. (Springer, 2016).
33. S. A. Adamo, "Bidirectional connections between the immune system and the nervous system in insects" in *Insect Immunology*, N. E. Beckage, Ed. (Elsevier, 2008).
34. S. G. Wubshet, J. S. Schmidt, S. Wiese, D. Staerk, High-resolution screening combined with HPLC-HRMS-SPE-NMR for identification of potential health-promoting constituents in Sea Aster and Searocket - New Nordic Food Ingredients. *J. Agric. Food Chem.* **61**, 8616–8623 (2013).
35. J.-H. Zhang, T. D. Y. Chung, K. R. Oldenburg, A simple statistical parameter for use in evaluation and validation of high throughput screening assays. J. Biomol. Screen. **4**, 67–73 (1999).
36. A. Birmingham, L. M. Selfors, T. Forster, D. Wrobel, J. C. Kennedy, E. Shanks, J. Santoyo-Lopez, D. J. Dunican, A. Long, D. Kelleher, Q. Smith, R. L. Beijersbergen, P. Ghazal, C. E. Shamu, Statistical methods for analysis of high-throughput RNA interference screens. *Nat. Methods* **6**, 569–575 (2009).
37. I. Protsyuk, A. Melnik, L.-F. Nothias, L. Rappez, P. Phapale, A. Aksenov, A. Bouslimani, S. Ryazanov, P. C. Dorrestein, T. Alexandrov, 3d molecular cartography using LC–MS facilitated by Optimus and 'ili software. *Nat. Protoc.* **13**, 134–154 (2018).
38. H. L. Röst, T. Sachsenberg, S. Aiche, C. Bielow, H. Weisser, F. Aicheler, S. Andreotti, H.-C. Ehrlich, P. Gutenbrunner, E. Kenar, X. Liang, S. Nahnsen, L. Nilse, J. Pfeuffer, G. Rosenberger, M. Rurik, U. Schmitt, J. Veit, M. Walzer, D. Wojnar, W. E. Wolski, O. Schilling, J. S. Choudhary, L. Malmström, R. Aebersold, K. Reinert, O. Kohlbacher, OpenMS: A flexible open-source software platform for mass spectrometry data analysis. *Nat. Methods* **13**, 741-748 (2016).
39. P. Shannon, A. Markiel, O. Ozier, N. S. Baliga, J. T. Wang, D. Ramage, N. Amin, B., Schwikowski, T. Ideker, Cytoscape: A software environment for integrated models of biomolecular interaction networks. *Genome Res.* **13**, 2498–2504 (2003).

40. T. Pluskal, S. Castillo, A. Villar-Briones, M. Orešič, MZmine 2: Modular framework for processing, visualizing, and analyzing mass spectrometry-based molecular profile data. *BMC Bioinformatics* **11**, 395 (2010).
41. P. Cignoni, M. Callieri, M. Corsini, M. Dellepiane, F. Ganovelli, G. Ranzuglia, MeshLab: an open-source mesh processing tool. *Sixth Eurographics Italian Chapter Conference*, 129-136 (2008).
42. A. Bouslimani, C. Porto, C. M. Rath, M. Wang, Y. Guo, A. Gonzalez, D. Berg-Lyon, G. Ackermann, G. J. Moeller Christensen, T. Nakatsuji, L. Zhang, A. W. Borkowski, M. J. Meehan, K. Dorrestein, R. L. Gallo, N. Bandeira, R. Knight, T. Alexandrov, P. C. Dorrestein, Molecular cartography of the human skin surface in 3D. *Proc. Natl. Acad. Sci. U S A* **112**, E2120–E2129 (2015).
43. B. E. Sedio, Recent breakthroughs in metabolomics promise to reveal the cryptic chemical traits that mediate plant community composition, character evolution and lineage diversification. *New Phytol.* **214**, 952–958 (2017).
44. O. L. M. da Silva, I. Cordeiro, *Euphorbia sarcoceras*, a new species of *Euphorbia* sect. Alectoroctonum from Brazil. *Syst. Bot.* **40**, 962–967 (2016).
45. Y. Yang, R. Riina, J. J. Morawetz, T. Haevermans, X. Aubriot, P. E. Berry, Molecular phylogenetics and classification of *Euphorbia* subgenus *Chamaesyce* (Euphorbiaceae). *TAXON* **61**, 764–789 (2012).
46. L. J. Revell, phytools: An R package for phylogenetic comparative biology (and other things). *Methods Ecol. Evol.* **3**, 217–223 (2012).
47. D. Orme, R. Freckleton, G. Thomas, T. Petzoldt, S. Fritz, N. Isaac, W. Pearse, caper: Comparative analyses of phylogenetics and evolution in R. R package version 0.5.2. (2013).
48. M. R. E. Symonds, S. P. Blomberg, "A Primer on phylogenetic generalised least squares" in *Modern Phylogenetic Comparative Methods and Their Application in Evolutionary Biology: Concepts and Practice*, L. Z. Garamszegi, Ed. (Springer, 2014).
49. R. Govaerts, F. J. F. Casas, C. Barker, S. Carter, S. Davies, H.-J. Esser, M. Gilbert, P. Hoffmann, A. Radcliffe-Smith, V. Steinmann, P. van Welzen, T. Whitmoore, World Checklist of Euphorbiaceae. Facilitated by the Royal Botanic Gardens, Kew, Published on the Internet; http://apps.kew.org/wcsp/, Retrieved 2014-07-25 (2014).
50. A. K. Hundsdoerfer, I. J. Kitching, M. Wink, The phylogeny of the *Hyles euphorbiae* complex (Lepidoptera: Sphingidae): Molecular evidence from sequence data and ISSR-PCR fingerprints. Org. Divers. Evol. **5**, 173 – 198 (2005).
51. J. C. Meerman, Relationships within the *Hyles euphorbiae* complex: a numerical taxonomy approach (Lepidoptera: Sphingidae). Entomol. Gaz. **44**, 205–209 (1993).

Acknowledgments: The authors thank Ivan Protsyuk and Theodore Alexandrov (EMBL) for their support in using the KNIME workflow Optimus and MassFuser, Arife Önder for technical assistance with fractionation, Martin Aarseth-Hansen and Gry Bastholm for assistance with collecting plant material and co-workers of LEO Pharma for technical assistance with biological assays. Kevin Nguyen, Brandon Ross and Kirandeep Chhokar (UCSD) are thanked for assistance with preparing plant extracts for the 3D mass spectral molecular cartography; **Funding:** This work was supported by the Marie Curie Actions of the 7th European Community Framework Programme: FP7/2007-2013/, REA grant agreement n 606895-MedPlant to NR and a grant from the Aase and Ejnar Danielsens Foundation to ME, and the MSCA-IF-2016 n 704786 to LFN.

The TNF-α experiments were conducted at and partly funded by LEO Pharma A/S. LC-MS/MS analyses were funded through the Center for Computational Mass Spectrometry P41 GM103484 and the NIH grant on reuse of metabolomics data R03 CA211211; **Author contributions:** ME, LFN, CHSL, OMG, HTS, NN, DS, PCD and NR designed the study. ME and KM collected and sampled *Euphorbia* species from the Botanical Garden at the Natural History Museum of Denmark. GH and LAF collected seeds of Brazilian *Euphorbia* species and performed taxonomic circumscription and identification. ME prepared the plant extracts. LFN performed LC-MS/MS analysis of the pooled *Euphorbia* extracts. ME performed LC-MS/MS analysis for the 3D mass spectral molecular cartography. ME and LFN performed mass spectral molecular networking analysis. JJJVDH performed MS2LDA analysis. LFN, JJJVDH and ME annotated Mass2Motifs on MS2LDA. ME and JJJVDH developed and performed the (semi-)automated annotation workflow by combining mass spectral molecular networking with MS2LA, *in silico* annotation and ClassyFire. RRS developed dereplication strategies through NAP and provided support for data analysis. NN, PL, DS and ME designed the high-resolution bioactivity study. ME performed the high-resolution bioactivity study and analysed data together with DS, NN and PL. ME generated the phylogenetic hypothesis, performed comparative analyses and wrote all scripts used for data analysis. ME wrote the manuscript together with LFN, NR and PCD. All authors discussed the results and commented on the manuscript. **Competing interests:** Authors declare no competing interests. **Data and materials availability:** All data, code, and materials to understand and assess the conclusions of this research are available in the main text, supplementary materials, and via https://github.com/DorresteinLaboratory/supplementary-GlobalEuphorbiaStudy. LC-MS/MS data are publicly accessible on GNPS under the MassIVE accession no. MSV000081082 and MSV000081083.

Supplementary Materials:

Materials and Methods

Supplementary Text

Table S1

Fig S1-S12

URL S1

Data S1-S2

References (*31-51*)

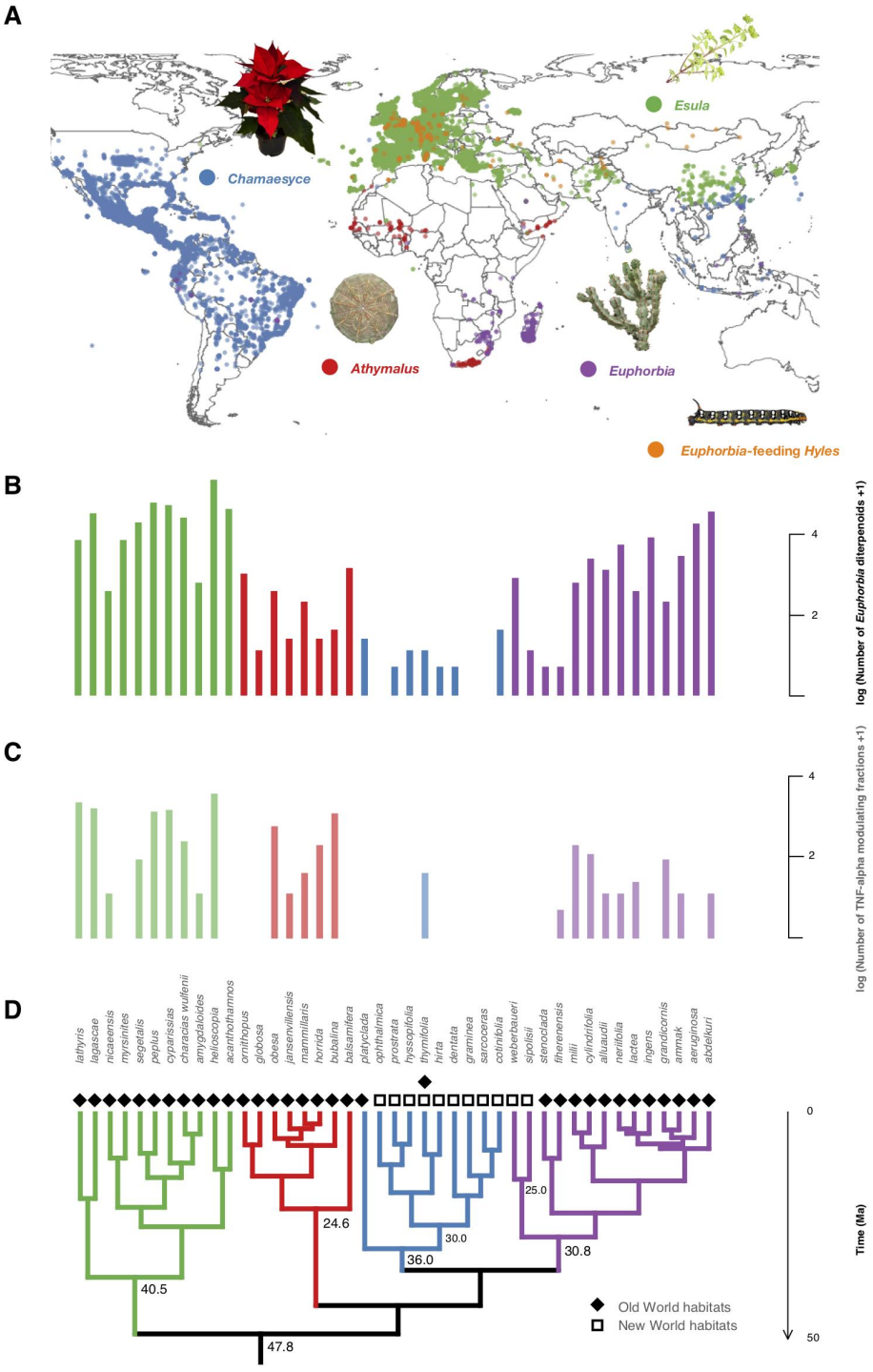

Fig. 1. Biogeography, phylogenetic relationships, diterpenoid production and biological activities of representative *Euphorbia* species. **A**. Occurrences of *Euphorbia* species investigated chemically and *Euphorbia*-feeding *Hyles* moth larvae retrieved from GBIF and manually restricted to native areas **B**. Number of putatively annotated *Euphorbia* diterpenoids per species analyzed. **C**. Number of TNF-α modulating fractions per species

analyzed. **D.** *Euphorbia* phylogenetic tree (50% majority rule consensus tree from Bayesian analysis of 11587 bps of DNA markers spanning all three plant genomes: chloroplast, mitochondrial, nuclear). Species of subgenus *Esula* exhibit a high number of biologically active diterpenoids and co-occur with larvae of *Euphorbia*-feeding *Hyles*, whereas the American radiation of subgenus *Chamaesyce* shows reduced *Euphorbia* diterpenoid production and TNF-α modulating activity. Subgeneric clades are highlighted with different colors: *Athymalus* (red), *Chamaesyce* (blue), *Esula* (green), *Euphorbia* (purple).

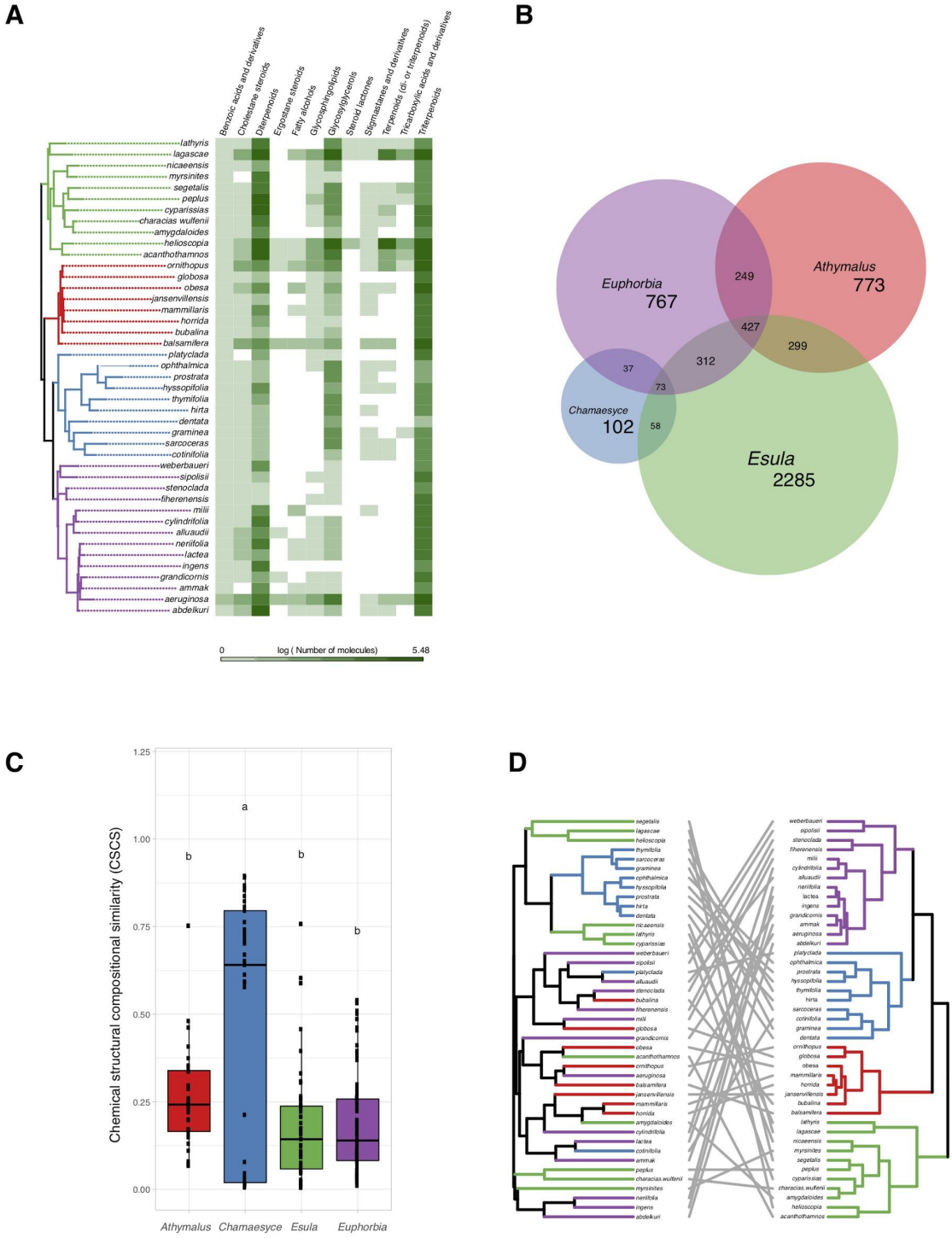

Fig. 2. Specialized metabolite diversity in *Euphorbia*. **A**. Distribution of specialized metabolite classes on the *Euphorbia* phylogenetic tree (50% majority rule consensus tree from Bayesian analysis of 11587 bps of DNA markers spanning all three plant genomes: chloroplast, mitochondrial, nuclear). Chemical classes of *Euphorbia* specialized metabolites were identified using a mass spectrometry based workflow combining mass spectral

molecular networking, *in silico* annotation, automated chemical classification and substructure recognition. **B.** Molecular features representing individual mass spectral molecular network nodes shared across species of *Euphorbia* subgeneric clades. **C**. Chemical similarity among *Euphorbia* subgeneric clades assessed using the chemical structural compositional similarity. Compared to subgenera *Athymalus*, *Esula* and *Euphorbia*, subgenus *Chamaesyce* exhibits very few chemically distinct features and high chemical structural compositional similarity. **D.** *Euphorbia* chemogram (left) and phylogenetic tree (right). The chemogram was generated using hierarchical cluster analysis on the pair-wise chemical structural and compositional dissimilarities of the tandem mass spectrometry data of the crude extracts using the complete agglomeration method. Phylogeny and chemogram show low overlap, suggesting that closely related *Euphorbia* species differ considerably in their chemistry.

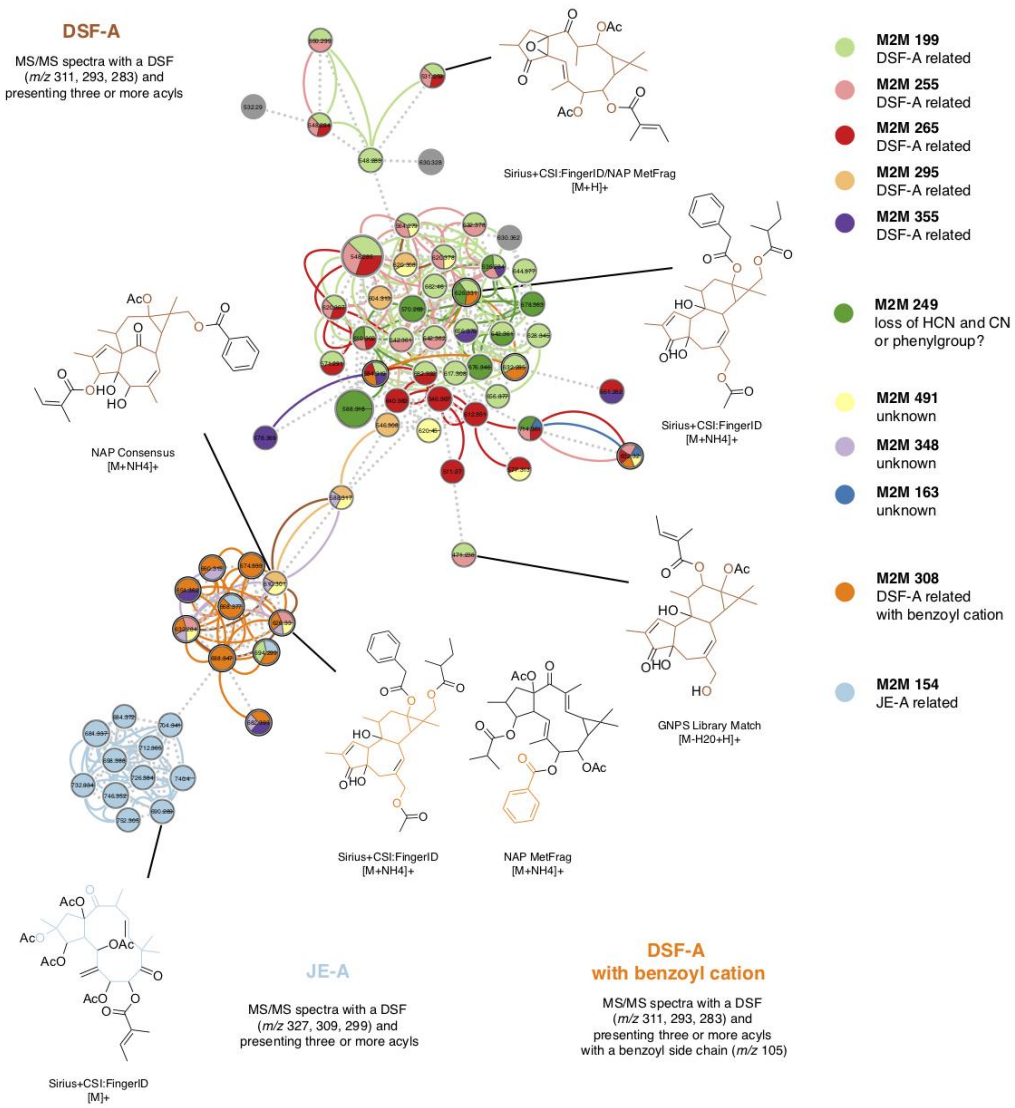

Fig. 3. Putative identification of chemical compound classes. We putatively identified compound classes within the mass spectral molecular networks by combining *in silico* annotation with automated chemical classification and substructure recognition (MS2LDA). *Euphorbia* diterpenoids exhibit many isoforms, therefore different diterpene backbone skeletons were found within the same molecular family. Matching substructures (Mass2Motifs) associated with diterpenoid substructures obtained from matches to reference spectra and *in silico* structure annotation enabled the identification of different diterpene spectral fingerprints clustered within one molecular family. Node size represents the total ion current (TIC) of all samples analyzed, edge colors represent different substructures (Mass2Motifs) that are shared across different nodes and dotted lines connecting the nodes represent the cosine score. M2M: Mass2Motif, DSF-A: Diterpene spectral fingerprint type A, JE-A: Jatrophane ester type A.

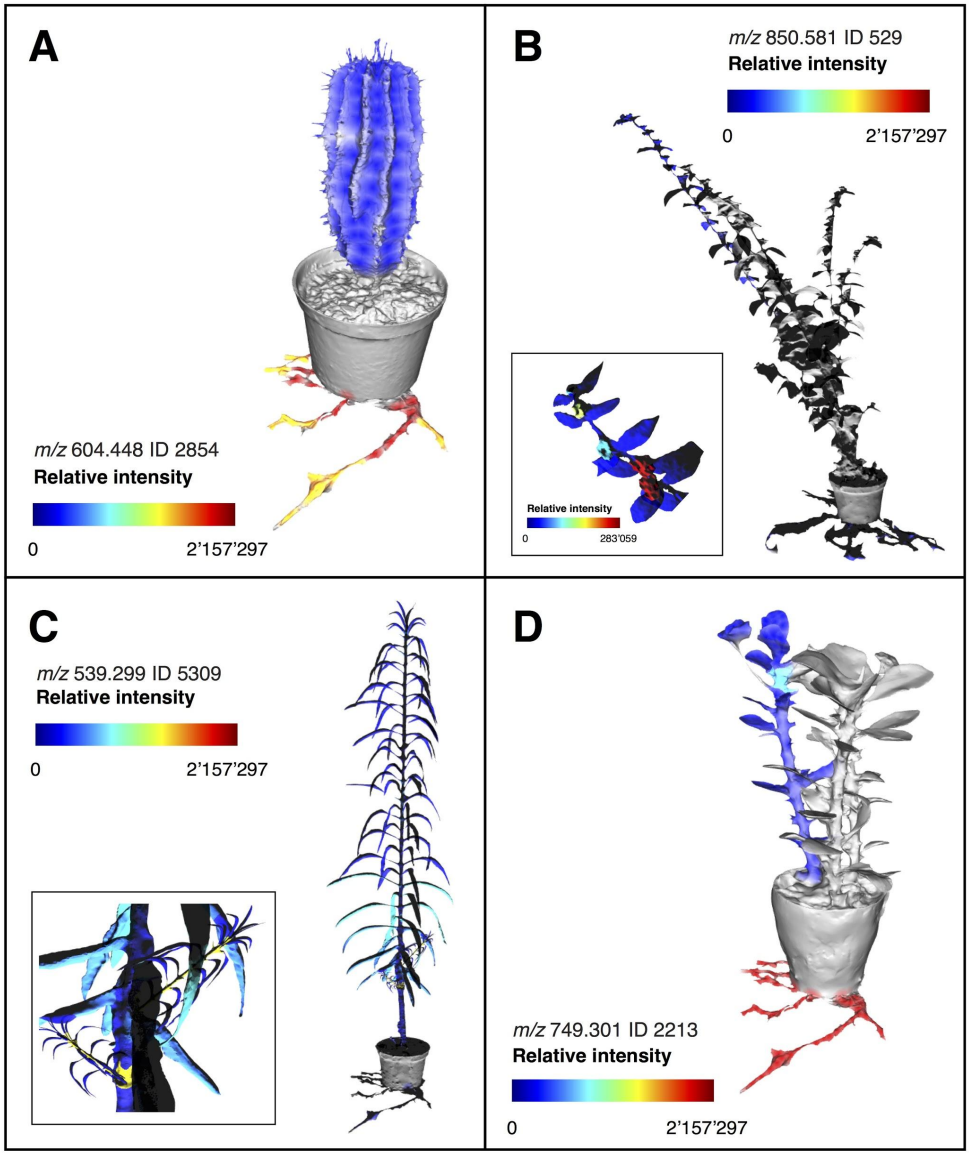

Fig. 4. Molecular maps of selected *Euphorbia* diterpenoids across representatives of each subgeneric clade. Relative intensity of LC-MS molecular features annotated as *Euphorbia* diterpenoids through spectral matching **A**. *Euphorbia horrida*, subgenus *Athymalus*, **B**. *Euphorbia hirta*, subgenus *Chamaesyce*, **C**. *Euphorbia lathyris*, subgenus *Esula* and **D**. *Euphorbia milii* var. *hislopii*, subgenus *Euphorbia*. For interactive cartographical snapshots see URL S1, links 1-15. The 3D images are for illustrative purposes only and do not represent exact locations of sample collection.