VIP *Very Important Paper*

Identification and Composition of Clasper Scent Gland Components of the Butterfly *Heliconius erato* and Its Relation to Mimicry

Stephanie Ehlers,[a] Daiane Szczerbowski,[a] Tim Harig,[a] Matthew Stell,[a] Susan Hötling,[a] Kathy Darragh,[b] Chris D. Jiggins,[c] and Stefan Schulz*[a]

Dedicated to the memory of Wittko Francke.

The butterfly *Heliconius erato* occurs in various mimetic morphs. The male clasper scent gland releases an anti-aphrodisiac pheromone and additionally contains a complex mixture of up to 350 components, varying between individuals. In 114 samples of five different mimicry groups and their hybrids 750 different compounds were detected by gas chromatography/mass spectrometry (GC/MS). Many unknown components occurred, which were identified using their mass spectra, gas chromatography/infrared spectroscopy (GC/IR)-analyses, derivatization, and synthesis. Key compounds proved to be various esters of 3-oxohexan-1-ol and (*Z*)-3-hexen-1-ol with (*S*)-2,3-dihydrofarnesoic acid, accompanied by a large variety of other esters with longer terpene acids, fatty acids, and various alcohols. In addition, linear terpenes with up to seven uniformly connected isoprene units occur, e.g. farnesylfarnesol. A large number of the compounds have not been reported before from nature. Discriminant analyses of principal components of the gland contents showed that the iridescent mimicry group differs strongly from the other, mostly also separated, mimicry groups. Comparison with data from other species indicated that *Heliconius* recruits different biosynthetic pathways in a species-specific manner for semiochemical formation.

Introduction

Heliconius (Nymphalidae) constitutes a species-rich genus of neotropical butterflies that show wide variations in colorful wing patterns within and between species. Bates described how the wing pattern of a single species would change with geographic location, and also observed convergence in wing pattern between different species at the same location.[1] *Heliconius* butterflies are unpalatable due to the presence of cyanogenic glycosides that can be either sequestered from *Passiflora* on which they feed or synthesized *de novo*.[2] Converging on the same warning pattern is beneficial because the butterflies can more efficiently advertise their unpalatability, a phenomenon known as Müllerian mimicry.[3]

The interesting mimicry situation, in which a species can carry location specific wing-patterns shared with other species and fast species evolution stirred the interest in the investigation of chemical signals released by the butterflies. Although, in contrast to moths, visual communication seems to be predominant in butterflies, many species additionally use volatile or lipophilic compounds to convey information.[4]

Different male pheromones have been described in *Heliconius*. Males possess a clasper scent gland (CSG) on their abdomen that contains a complex mixture of compounds. These compounds are transmitted during copulation to the female where they serve as anti-aphrodisiac pheromone,[5,6] inhibiting courtship of other males and ensuring undisturbed egg-laying of the female. Males additionally carry areas covered with scent disseminating scales on their wings, so called androconia, which emit a pheromone affecting the behavior of females.[7,8] The chemical nature of the pheromones originating from the CSG and the androconia is different (Figure 1).

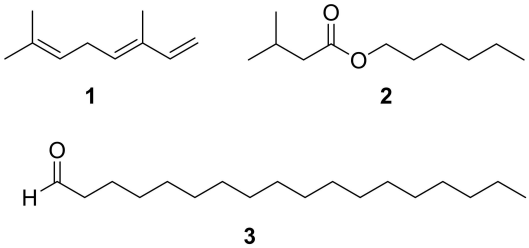


Figure 1. Known pheromones of *H. melpomene*, (*E*)-β-ocimene (**1**), octadecanal (**3**), as well as hexyl 3-methybutanoate (**2**) of *H. cydno*.

[a] *S. Ehlers, Dr. D. Szczerbowski, Dr. T. Harig, Dr. M. Stell, Dr. S. Hötling, Prof. Dr. S. Schulz*
Technische Universität Braunschweig
Hagenring 30, 38106 Braunschweig (Germany)
E-mail: stefan.schulz@tu-bs.de
Homepage: http://www.oc.tu-bs.de/schulz/index_en.html
[b] *Dr. K. Darragh*
Department of Evolution and Ecology,
Storer Hall University of California
One Shields Avenue, Davis CA, 95616 (USA)
[c] *Prof. Dr. C. D. Jiggins*
Department of Zoology, University of Cambridge
Downing Street, CB2 3EJ Cambridge (UK)

Supporting information for this article is available on the WWW under https://doi.org/10.1002/cbic.202100372

Heliconius melpomene produces large amounts of (*E*)-β-ocimene (**1**) in the CSG, accompanied with a complex mixture of less volatile, mostly aliphatic esters. Volatile **1** functions as an anti-aphrodisiac whereas the esters seem to modify the evaporation rate.[6] Hexyl 3-methybutanoate (**2**) has the same function in *H. cydno*.[9] In contrast, the androconia of *H. melpomene* release octadecanal, much less volatile compared to **1**, serving as a courtship signal.[7,10]

The CSG and androconial compound composition has been analyzed in a range of heliconiines.[8,10–15] Generally, androconia seem to contain few components, often being long chain aldehydes, alcohols, acetates, hydrocarbons and few likely lignin-derived aromatic compounds such as syringaldehyde.[10] In contrast, the CSG gland content is chemically more diverse with a large number of compounds of varying size and biosynthetic origin as well as species-specific mixtures, many of them not yet characterized. Usually volatile compounds like **1** are present, but accompanied by much larger compounds such as fatty acid esters that have been assumed to function as matrix for a volatile anti-aphrodisiac signal.[6] Nevertheless, a high species-specificity within the larger compounds advocates a more specific role of these compounds in the chemical communication of the butterflies.

Heliconius erato occurs in various subspecies with different mimetic wing color patterns throughout Central and South America. The forms relevant for our study are shown in Figure 2. *Heliconius erato* contains a particular abundance of different compounds in the CSG, but also shows intraspecific variation in its gland contents. In a recent study we analyzed 104 individuals and showed distinct differences in CSG gland composition between *H. erato* from West Ecuador and other localities.[13] Although about 230 compounds were detected, 50% of them remained structurally uncharacterized, some of them major components of the secretion. During that study and from earlier analyses[12] it became clear that many of these compounds occur in other *Heliconius* species as well, differing often in concentration. Therefore, an effort was made to structurally identify as many metabolites of the *H. erato* CSG as possible. The knowledge of compound structures allows us to understand their biosynthetic origin and their distribution within the heliconiines, and helps to understand their function. It sets the basis for functional characterization of the CSG secretion. Alongside the discovery of new natural products, comparative metabolomic analysis will greatly benefit from the knowledge of compound identities, showing the need for correct compound characterization in metabolomic analyses.

Here we report an in-depth analysis of the composition of the *H. erato* CSG compounds by gas chromatography/mass spectrometry (GC/MS) and gas chromatography/infrared spectroscopy (GC/IR) including their structural identification, backed up by the synthesis of key components for structural proof and for later tests of biological activity. Furthermore, we will discuss the results in the context of mimicry and compositional variation. The biosynthetic background in light of the recently identified ocimene synthase from *H. melpomene*[16] and the available *Heliconius* genome[17] will be discussed. Specifically, we want to address whether wing pattern mimicry is reflected in CSG composition and whether specific composition variation can be found in different volatility ranges of the odor bouquet. Furthermore, using our previous data[13], updated here, a comparative analysis indicates different expression of basic biosynthetic pathways within the heliconiines.

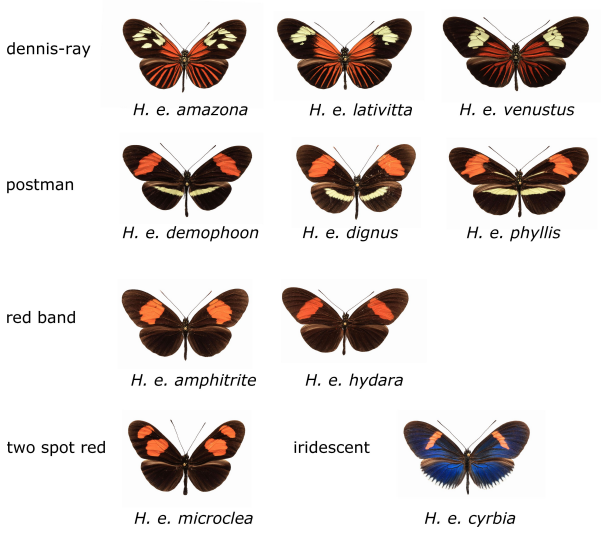


Figure 2. Wing patterns of the different mimicry groups of *H. erato* subspecies investigated. *Heliconius erato amazona* (Brazil), *H. erato lativitta* (eastern Ecuador and Colombia), *H. erato venustus* (Brazil), *H. erato demophoon* (Panama), *H. erato dignus* (Colombia), *H. erato phyllis* (Brazil), *H. erato amphitrite* (Peru), *H. erato hydara* (Brazil), *H. erato microclea* (Peru), and *H. erato cyrbia* (Western Ecuador).

Results and Discussion

In total, 114 samples including 10 subspecies and different hybrids collected from different locations in Central and South America were analyzed by GC/MS. Most of the samples had been analyzed before with a cutoff gas chromatographic retention indices (*I*) of >2900, but many compounds remained structurally uncharacterized and trace components were excluded.[13] Identification was performed with commercial and in-house databases and compound identity was assured by synthesis, if necessary. In summary, 750 different compounds were detected without contaminants (see SI, Experimental part), and 550, occurring at least in three samples were included in further analyses. Thus, it was ensured that all compounds occurring at least in half of the number of the samples of a mimicry group were included. After completion of our analysis, most of the still unidentified compounds occurred only in a few specimens.

Generally, several structural motifs occurred. Main compounds were terpenes and their esters with short chain alcohols. The latter were also esterified to fatty acids of various chain length. In addition, various other compounds such as aromatic compounds occurred. The total ion chromatogram (TIC) often showed a typical distribution of compounds (Figure 3). In the early part usually (*E*)-β-ocimene (**1**) and 3-

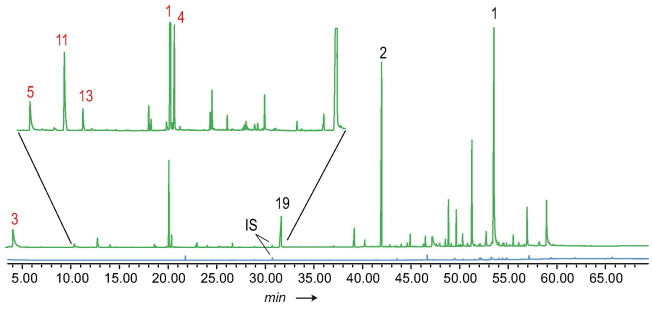


Figure 3. Total ion chromatogram (TIC) of *Heliconius erato lativitta* from Ecuador. CSG extract (green) and wing area extract (blue) containing the same amount of internal standard (IS). Black numbers assign compounds using entry numbers of Table 1, red ones of Table 4.

undecanone (**39**) are dominant compounds. There follows a region with few components, whereas later in the chromatogram a large number of compounds occurs. Nevertheless, large differences between samples were found (see Figures S1-S6 in the SI for examples). Because an apolar GC phase was used, the retention times follow roughly the boiling points of the compounds and the late eluting compounds were previously referred to as matrix to modulate evaporation of the volatile signal **1** after transfer to the female.[6,11] Nevertheless, the high structural diversity found among these compounds within the heliconiines might indicate additional, currently unknown functions, probably acting as close range signals.

Whereas the structure of **1** and **39** was confirmed by comparison with a synthetic sample,[18] many other compounds were structurally unknown. The identification will be described in the following paragraphs.

Compound identification

Three late eluting components **A**/**A'** and **B** (see Figure S1 in the SI) exhibited the mass spectra shown in Figure 4. The spectra showed similarities with the spectrum of 2,3-dihydrofarnesoic acid (DHF, **8a**), which was also present in the secretion and proved to be the (*E*)-diastereomer by synthesis (Scheme 1). Both enantiomers of **8a** were synthesized from (2*E*,6*E*)-farnesol (**4**) by enantioselective hydrogenation using Noyori's method,[19] followed by Parikh-Döring and Pinninck-Lindgren oxidations[20] to arrive at **8a**.

The mass spectrum of **8a** is characterized by the ions *m/z* 69, 109, and 123 (Figure 4). Whereas *m/z* 69 is formed by allylic cleavage (Figure 5), the other two ions require rearrangements as explained in the supplementary information (Figure S7, SI). Of special importance is the ion M-43, *m/z* 195. As has been shown by deuterium labeling,[21] long chain esters can lose C-2 and C-3 and an additional H, especially when being methyl substituted at these positions. A possible explanation for this unusual loss of an odd-numbered fragment from within the chain is shown in Figure 5. Radical cation **10** induces rearrangement to **11** by C-3–C-4 bond cleavage and transfer to O. A C-3 substituent enhances this pathway due to the formation of a

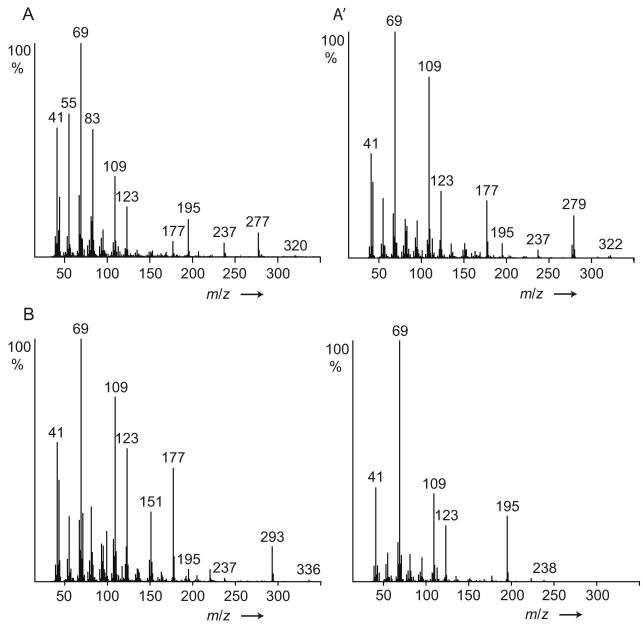
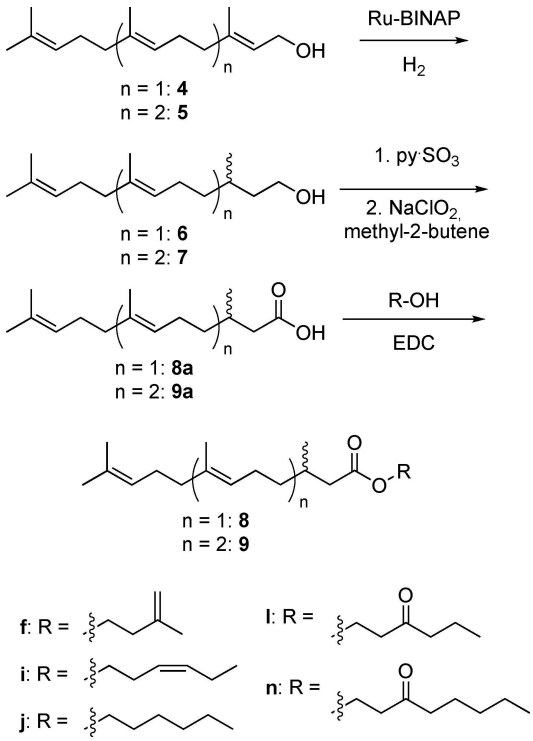


Figure 4. Mass spectra of (*E*)-2,3-dihydrofarnesoic acid (**8a**, lower right), and unknown compounds **A**, **A'** and **B**, identified as esters **8i**, **8j**, and **8l**, respectively.



Scheme 1. Enantioselective synthesis of (*E*)-2,3-dihydrofarnesoic acid (**8a**), geranylcitronellic acid (**9a**) and its esters. Various enantiomers were synthesized.

secondary radical. Distonic radical cation **10** can now abstract an H from C-5, leading to the allylic radical **12**. Finally, α-cleavage and ring closure leads to the relatively stable cation

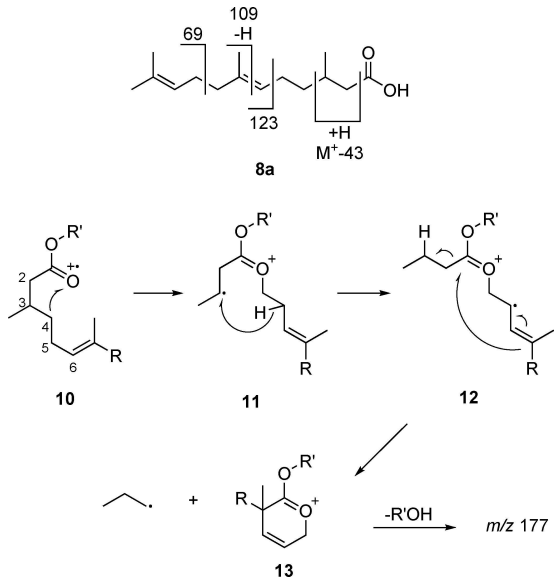


Figure 5. Proposed mass spectral fragmentation of 2,3-dihydrofarnesoic acid (**8a**) and its esters.

m/z 197 (**13**). The high abundance of this ion can therefore be explained not only due to the methyl group at C-3, but also by the presence of the C-6 double bond.

In summary, abundant ions *m/z* 69, 109, 123 and M^+-43 are good indicators of compounds containing the 2,3-dihydrofarnesyl system, all found in **A/A'** and **B**. The compound pair **A/A'** showed in addition M^+-ions at *m/z* 320/322, suggesting that they are hexyl and hexenyl esters. (*Z*)-3-Hexenol, the leaf alcohol, is a common plant constituent formed by lipid degradation, e.g. induced by insect feeding. In addition, an IR spectrum of **A/A'**, obtained by GC/direct deposition infrared spectroscopy (GC/DD-IR, see SI Figure S26), not separable under the conditions, showed a band characteristic for a disubstituted (*Z*)-double bond at 3009 cm^{-1}.[22] Both compounds, hexyl (*E*)-2,3-dihydrofarnesoate (**8j**) and (*Z*)-3-hexenyl (*E*)-2,3-dihydrofarnesoate (**8i**) were synthesized by esterification of the respective alcohol with **8a** using EDC (Scheme 1). Compound **B** showed a molecular weight 14 amu higher than **8i**. The GC/DD-IR-spectrum (Figure S27, SI) showed a broadened CO-ester band with a shoulder below 1734 cm^{-1}, indicating an additional keto-group. Furthermore, the mass spectrum exhibited ions at *m/z* 71 and 99, both consistent with 3-oxohexanol as alcohol part of **B**. Therefore, 3-oxohexyl (*E*)-2,3-dihydrofarnesoate (**8l**) was synthesized from 3-oxohexanol (see SI Scheme S1) and confirmed as the structure of **B**. The mass spectra of the esters show in addition to the ions *m/z* 69, 109, 123 fragment *m/z* 177, likely formed by loss of ROH from **13**. These ions now allowed to search the GC/MS data of the various samples for other dihydrofarnesyl esters, leading to the identification of ethyl (**8c**), propyl (**8d**), butyl (**8e**), 3-methyl-2-butenyl (**8g**, prenyl), 3-oxoheptyl (**8m**), methyl (**8b**), 3-methylbutyl (**8h**), 3-methyl-3-butenyl (**8f**, isoprenyl), and 3-oxooctyl (**8n**) esters (Figure 6), of which the last four were synthesized. A dihydrofarnesyl (DHF) ester in which the ions *m/z* 71 and 99 were shifted to *m/z* 69

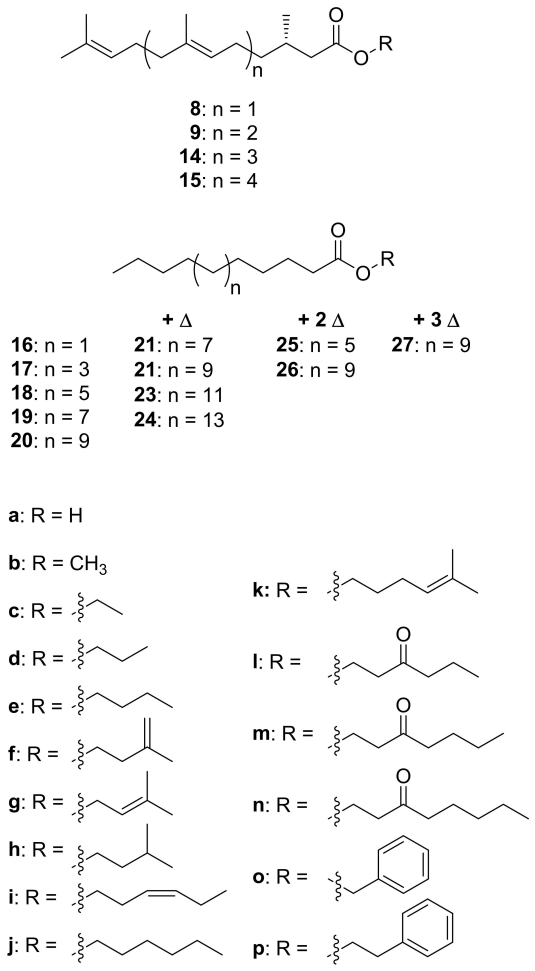


Figure 6. Esters occurring in the CSG secretion of *H. erato*.

and 97 indicated a 3-oxohexenyl residue, whereas an intense ion *m/z* 105 besides the discussed characteristic DHF fragments showed the presence of 2-phenylethyl (*E*)-2,3-dihydrofarnesoate (**8p**), whereas *m/z* 91 indicated benzyl (*E*)-2,3-dihydrofarnesoate (**8o**). Mass spectra of esters are shown in the SI.

The direct determination of the absolute configuration of the DHF esters by enantioselective GC was not possible due to the high elution temperatures needed. Therefore, transesterification into the methyl ester **8b** was performed with trimethylsulfonium hydroxide[23] because the synthesized enantiomers proved to be separable. In all cases investigated, *H. erato lativitta*, *venustus*, *microclea* and *emma* hybrids, only the (*S*)-enantiomer was detected, verifying its exclusive formation in various subspecies (Figure 7 and Figure S37, SI).

All the DHF esters have not been reported from nature before except methyl (*E,R*)-2,3-dihydrofarnesoate (*R*-**8b**), which is part of the pheromone of various *Chlorochroa* sting bugs,[24] whereas acid *S*-**8a** occurs in the cephalic secretion of male European beewolves *Philanthus triangulum*,[25] as well as in tomato trichomes.[26] Short chain alcohols present in the esters can be biosynthetically formed from the primary metabolism. Isoprenyl and 3-methylbutyl alcohols are likely derived from the

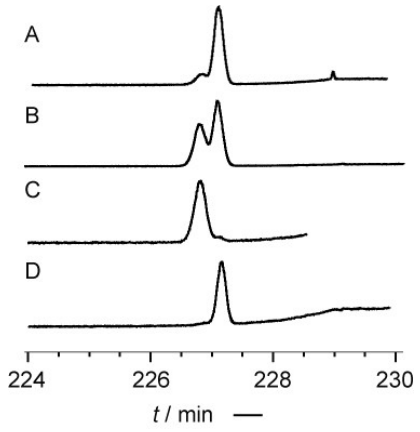


Figure 7. Enantiomer separation of methyl (*E*)-2,3-dihydrofarnesoate (**8b**) from *H. erato lativitta* by enantioselective GC on a heptakis-(2,3-di-*O*-methyl-6-*O*-*tert*-butyldimethylsilyl)-β-cyclodextrin (Hydrodex-6-TBDMS) phase. A: Co-injection of racemic **8b** and natural extract; B: *S*- and *R*-**8b**; C: *R*-**8b**; D: *S*-**8b**.

terpene building block isopentenyl pyrophosphate. (*Z*)-3-Hexenol, the leaf alcohol, its oxidation product 3-oxohexanol, and 1-hexanol are products of chain oxidation and cleavage of unsaturated fatty acids.[27]

3-Oxohexanol have so far been reported only as a component in 3-oxohexyl ferulate from *Baccharis*[28] and *Larrea tridentata* plants[29] and thus represents a unique feature of the *H. erato* CSG bouquet. The characteristic MS behavior with the ion doublet *m/z* 71/99 served as tool to detect further 3-oxohexyl esters. Compound no. 16 in Table 1 showed such a mass spectrum with additional ions similar to those found in the DHF esters, and a molecular mass of 404 D, 68 D higher than **8l** (Figure S16). We proposed, therefore, this compound to be 3-oxohexyl geranylcitronellate (**9l**). To prove this proposal, (*E*,*E*)-geranylgeraniol, obtained from annatto seeds,[30] was enantioselectively hydrogenated and transformed analogously to the DHF esters into the respective hexyl, (*Z*)-3-hexenyl and 3-oxohexyl esters **9i**, **9j** and **9l**. All three compounds were detected in the CSG secretions. Finally, unusual higher [5]- and [6]-terpene esters[1] such as (*Z*)-3-hexenyl 2,3-dihydrogeranylfarnesoate (**14i**) and (*Z*)-3-hexenyl 2,3-dihydrofarnesylfarnesoate (**15i**) were identified using the described methods including transesterification to the respective methyl esters **14b** and **15b**.

Whereas the esters **8i,j,l** usually dominated the CSG secretion, fatty acid esters with many of the residues **a**–**p** were also detected, albeit in lower concentration (Table 1). Acids occurring in such esters were a bishomologues series of saturated acids from C_{10} to $C_{20,}$ and unsaturated analogs from C_{14} to C_{22}. Dimethyl disulfide derivatization of methyl octadecenoate[32] obtained from transesterification showed the double bond to be located at C-9, consistent with previous results from androconial secretion analysis of *H. melpomene*.[10] Similar esters are known as constituents of CSG glands from other heliconiines,[6,12,13] but have also been reported from other butterfly genera such as the ithomiine *Ithomia*,[33] the nymphalid *Bicyclus*,[34] or the pierids C*olias*[35] and *Pieris*.[36] Bioassays and electrophysiology showed that hexyl tetradecanoate (**18j**) likely acts as a species recognition signal in *C. eurytheme* and has together with other hexyl esters an aphrodisiac function.[35] 3-Oxohexyl fatty acid esters are specific to *H. erato* and can be identified due to their characteristic mass spectra. The mass spectrum of 3-oxohexyl hexadecanoate (**19l**, see SI Figure S19) features a base peak at *m/z* 71 with a prominent *m/z* 98/99 ion pair, together with the McLafferty-ion *m/z* 158, the acylium-ion *m/z* 239, and M^+ *m/z* 326. These make respective esters with

Table 1. Compounds occurring in highest mean amounts in CSG extracts (Top 50). The average value were calculated excluding samples that did not contain the compound. *I*: gas chromatographic linear retention index; oc: number of occurrence in the 114 samples analyzed; µg: mean amount of the compound in CSG extracts containing them. A minimum of oc = 3 was required for listing. A full list of compounds can be found in the SI, Table S7.

	Compound	*I*	oc	µg
1	(2*E*,6*E*,10*E*,14*E*,18*E*)-Farnesylfarnesol (**30**)	3122	100	27.63
2	3-Oxohexyl (*E*)-2,3-dihydrofarnesoate (**8l**)	2312	113	13.44
3	(2*E*,6*E*,10*E*,14*E*,18*E*,22*E*)-Geranylgeranylfarnesol	3583	96	7.19
4	(6*E*,10*E*,14*E*)-β-Farnesylfarnesene (**28**)	2828	99	7.19
5	(2*E*,6*E*,10*E*,14*E*)-Geranylfarnesol	2646	99	6.24
6	3-Undecanone (**39**)	1283	99	5.65
7	(2*E*,6*E*,10*E*,14*E*,18*E*)-α-Farnesylfarnesene	2874	86	3.96
8	2-Phenylethyl tetradecanoate (**18p**)	2450	27	3.46
9	Unknown B135_7	2965	24	3.29
10	Unknown B69_2	2488	11	3.11
11	(*E*)-β-Ocimene (**1**)	1043	115	3.10
12	2-Phenylethyl dodecanoate (**17p**)	2243	13	2.76
13	Hexyl (*E*)-2,3-dihydrofarnesoate (**8j**)	2153	108	2.65
14	1-Hexen-3-one	761	3	2.56
15	(6*E*,10*E*,14*E*,18*E*,22*E*)-β-Geranylgeranylfarnesene	3289	98	2.08
16	3-Oxohexyl geranylcitronellate (**9l**)	2766	111	1.66
17	Unknown B135 (6)	2953	10	1.63
18	2-Phenylethyl octadecenoate_2 (**20p**)	2852	16	1.53
19	(*E*)-2,3-Dihydrofarnesoic acid (**8a**)	1763	92	1.43
20	2-Phenylethyl (*E*)-2,3-dihydrofarnesoate (**8p**)	2435	111	1.41
21	3-Oxooctyl (*E*)-2,3-dihydrofarnesoate (**8n**)	2497	111	1.31
22	Cholesterol	3080	116	1.15
23	Unknown B116_3	2900	13	1.10
24	(2*E*,6*E*,10*E*,14*E*,18*E*,22*E*)-α-Geranylgeranylfarnesene	3335	83	1.03
25	Unknown B116_1	2512	13	1.03
26	Isoprenyl (*E*)-2,3-dihydrogeranylfarnesoate (**15f**)	2939	79	1.02
27	Tetracosenolide	2950	79	0.99
28	Hexyl tetradecanoate (**18j**)	2172	26	0.98
29	Hexyl dodecanoate (**17j**)	1971	13	0.89
30	3-Oxohexyl tetradecanoate (**18l**)	2325	34	0.87
31	Hexacosen-1-ol	2861	105	0.84
32	Hexyl 9-octadecenoate (**22j**) and (*Z*)-3-Hexenyl 9-octadecenoate (**22i**)	2542	101	0.82
33	2-Phenylethyl octadecenoate_1 (**20p**)	2838	83	0.82
34	1-Tetracosanol	2683	101	0.78
35	(*Z*)-3-Hexenyl (*E*)-2,3-dihydrofarnesoate (**8i**)	2149	99	0.74
36	Tetracosen-1-ol	2657	112	0.73
37	(6*E*,10*E*)-*β*-Geranylfarnesene	2367	69	0.72
38	Hexyl (*E*)-2,3-dihydrogeranylfarnesoate (**14j**)	3061	87	0.71
39	Unknown B135_5	2943	39	0.69
40	Geranylcitronellic acid (**9a**)	2213	46	0.68
41	Unknown B135_12	3410	60	0.67
42	(2*E*,6*E*,10*E*,14*E*)-Geranylfarnesyl acetate	2745	84	0.66
43	(2-Nitroethyl)-benzene (**41**)	1281	101	0.66
44	Unknown octadecatrienoate	2844	13	0.66
45	(6*E*,10*E*,14*E*,18*E*)-Farnesylnerolidol (**31**)	2932	63	0.65
46	2-Phenylethyl hexadecanoate (**19p**)	2659	86	0.63
47	Unknown (*E*)-2,3-dihydrofarnesylfarnesoate_2	3513	31	0.58
48	3-Oxohexyl 9-octadecenoate (**22l**)	2703	89	0.54
49	(6*E*,10*E*,14*E*,18*E*,22*E*)-Geranylgeranylnerolidol	3406	46	0.54
50	Octacosenol	3067	47	0.52

[1] [6]-terpene: indicates the number of consecutively arranged isoprene units within the terpene backbone. As a more complex example, squalene is a [3^1+3^1]-terpene. See Ref. [31] for a detailed description.

other acyl groups easy to detect. Esters **17l** and **19l** were synthesized as representative members of this ester series. Fatty acid esters of the other alcohols (types **a**–**j**, **l**–**p**) were also present; their identification has been reported previously[6,33] or is analogous to 3-oxohexyl esters (**m**, **n**). Of special interest are 5-methyl-4-hexenyl esters (**k**) occurring in small amounts and which have not been described before. These esters show dominant ions at *m/z* 81 and 96 in their mass spectra, as well as a small acylium ion and an almost absent M^+ (Figure S23). The high intensity of the two ions can be explained by a favorable neutral loss of the acid part leading to a stable five membered ring, followed by a loss of CH_3 to form an allylic cation (Figure S8). The structure of 5-hexen-4-enyl 9-octadecenoate (**22k**) was proven by synthesis as shown in Figure S48 in the SI. Biosynthetically these **k**-type esters might be formed by elongation of dimethylallyl pyrophosphate building block, elongated by malonyl- or acetyl-CoA, followed by reduction to the alcohol and finally esterification. 5-Hepten-4-enyl esters have not been reported before as natural products.

In summary, the combinatorial nature of esters and their thus resulting high structural variability explains to a large degree the differences in CSG gland composition observed within the different *H. erato* subspecies and within individuals. The ester diversity also differentiates this species from other heliconiines with usually lower ester diversity. The terpene esters occur in higher abundance compared to the fatty acid esters.

Beside the esters, a different type of compounds occurred in the CSG, exemplified by the unknown compounds **C**–**F** (*I* 2844, 2893, 2966, 3132 respectively, mass spectra see Figures S20–S22 in the SI). Mass spectral database searches revealed similarity with the diterpenes α/β-springene, geranylgeraniol, and geranyllinalool, but the highest detectable ion was *m/z* 408 in **C** and **D** instead of 272 for the former compounds. The spectral similarity as well as the mass difference of 2×68 amu indicated higher springene homologues carrying two additional isoprene units in the chain in a linear arrangement of six prenyl units ([6]-terpenes) with farnesylfarnesene (**28**, **29**), farnesylfarnesol (**30**) and farnesylnerolidol (**31**) as structures (Figure 8). Whereas the mass spectra did not allow easy discrimination of the compounds, GC/DD-IR (DD: direct deposition) proved to be suitable. All compounds differ only in their head groups, carrying either an alcohol or an α- or β-type diene system. Compound **C** showed no OH-band, but bands at 3090 cm^{-1}, 3056 cm^{-1} and 1596 cm^{-1}, also present in myrcene (Figure S29). Therefore, compound **C** was identified to be β-farnesylfarnesene (**28**). In contrast, compound **D** showed bands at 3091 cm^{-1} as well as a doublet at 1641 cm^{-1} and 1607 cm^{-1} as does model compound **1**, showing absorbance at 3091 cm^{-1}, 1642 cm^{-1} and 1607 cm^{-1}, indicating α-farnesylfarnesene (**29**) as structure for **D**.

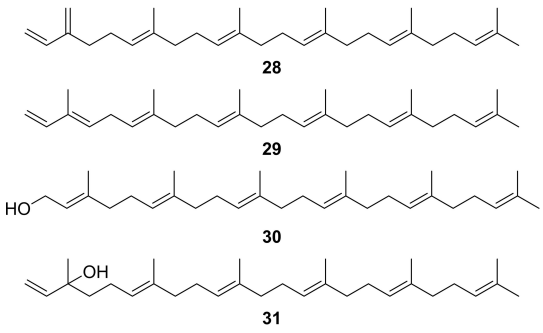


Figure 8. The linear [6]-terpenes (*E,E,E,E*)-β-farnesylfarnesene (**28**), (*E,E,E,E,E*)-α-farnesylfarnesene (**29**), (*E,E,E,E,E*)-farnesylfarnesol (**30**), and (*E,E,E,E*)-farnesylnerolidol (**31**).

Careful analysis of the extracts revealed the presence of higher and lower isoprenyl homologues of **28** and **29**, all showing a linear arrangement of isoprene units. These compounds occurred always in pairs with differences only in their head groups, being of the α- or β-type. The compounds and their respective gas chromatographic retention indices *I* are shown in Table 2. The identity of the [2]-[4]-terpenes was confirmed by comparison with synthetic samples from our compound library. The Δ*I* between the values within a column of Table 2 was always around 460, consistent with an isoprenyl (C_5H_8)- increment. A (*Z*)-double bond in any position would disrupt these values. Therefore, we can assign all-*E*-geometry for all compounds in this table. In addition, we synthesized (*E,E,E,E*)-β-geranylfarnesene from (*E,E,E*)-geranylgeraniol as described in the Supporting Information via bromination with PBr_3 and elongation with isoprenyllithium,[37] thus confirming our assignment.

Whereas we could not get a clean IR spectrum of **E** because of coeluting peaks, compound **F** showed a broad OH-band in its DD-IR spectrum. The IR spectrum was similar to that of synthetic (*E,E,E*)-geranylgeraniol (**5**) (Figure S34, SI), whereas it differed compared to (*E,E*)-geranyllinalool (Figure S35, SI) in the fingerprint region, missing for example the band at 921 cm^{-1}. Therefore, compound **F** was identified as (*E,E,E,E,E*)-farnesylfarnesol (**30**), whereas **E** was assigned to be (*E,E,E,E*)-farnesylnerolidol (**31**). These assignments were verified by *I* data, shown in

Table 2. Gas chromatographic retention indices *I* of hydrocarbon terpenes.

terpene	R	*I*	R	*I*
[2]	(*Z*/*E*)-ocimene	1039/1054	myrcene	990
[3]	(*E,E*)-α-farnesene (**50**)	1508	(*E*)-β-farnesene (**51**)	1455
[4]	(*E,E,E*)-α-springene (**53**)	1965	(*E,E*)-β-springene	1915
[5]	(*E,E,E,E*)-α-geranylfarnesene (**55**)	2430	(*E,E,E*)-β-geranylfarnesene (**56**)	2381
[6]	(*E,E,E,E,E*)-α-farnesylfarnesene (**29**)	2893	(*E,E,E,E*)-β-farnesylfarnesene (**28**)	2844
[7]	(*E,E,E,E,E,E*)-α-geranylgeranylfarnesene (**60**)	3357	(*E,E,E,E,E*)-β-geranylgeranylfarnesene (**59**)	3308

Table 3. Whereas isoprenyl unit homologues differed by 470 units, the difference between the farnesol and the nerolidol type was always around 170 units. Geometric isomers of these compounds are present as minor components, but on an apolar GC-phase the major all-*E* isomers always elute last, as known e.g. from farnesene and farnesol.[38] In summary, a suite of linear homologous terpene hydrocarbons and alcohols were present in the CSG secretion, ranging up to C_{35} or [7]-terpenes.

Only few of the [5]-[7]-terpenes have been reported before as natural products. (*E*,*E*,*E*)-β-geranylfarnesene (**56**) and (*E*,*E*,*E*,*E*)-β-farnesylfarnesene (**28**) have been reported from *Bacillus clausii*,[39] whereas the α-type hydrocarbons have not been reported before. (*E*,*E*,*E*,*E*)-Geranylfarnesol (**58**) is a component of the wax scale *Ceroplastes albolineatus*,[40] whereas (*E*,*E*,*E*)-geranylnerolidol (**57**) occurs in the phytopathogenic fungus *Cochliobolus heterostrophus*.[41] Farnesylfarnesols with unknown or different double bond configuration have been reported from some plants.[42]

The large diversity of CSG compounds with a relatively high molecular weight is due to the terpene variability in chain extension and the combinatorial ester formation of various acids with diverse alcohols. In contrast, only a limited number of more volatile compounds are present. These include short chain esters of 3-methylbutyric and 3-methyl-2-butenoic acids with the C_6-alcohols 3-oxohexanol, hexanol, and (*Z*)-3-hexenol (**32**–**36**, Figure 9). The esters 3-oxohexyl 3-methyl-2-butenoate, 3-methylbutanoate and dodecanoate were synthesized to confirm their identifications (see SI). In low amounts also 2-phenylethyl esters of these acids occur, but other compound classes are present as well. Major volatile compounds are ethyl- and vinylketones such as 3-undecanone (**39**), 1-octen-3-one (**37**), and 3-nonanone as well as related alcohols such as 3-undecanol (**38**). Occasionally, 1-hydroxy-3-hexanone (**40**) occurs. Mono- and sesquiterpenes are present, especially the anti-aphrodisiac **1**,[6] accompanied sometimes by its homologue, (*E*,*E*)-α-farnesene (**50**), δ-cadinene (**44**), or T-cadinol (**45**). Another type of terpenes, apocarotenoids,[43] are likely formed by degradation of probably diet-ingested carotenoids, such as dihydro-β-ionone (**47**), β-ionone (**46**) and dihydroedulan II (**48**).[44,45] For example, **48** has been reported first as a trace component of *Passiflora edulis*, the larval host of *Heliconius*.[45] The fungal volatile benzopyranone mellein (**43**) is known from other butterflies such as danaids[46] or lycaenids[47] as well as from ants where it functions for example as a trail pheromone.[48] The occurrence of **43** in fungi, plants and insects as well as bacteria[49] leaves its origin unclear, be it by uptake, de novo biosynthesis, or even symbiosis with microorganisms. Of particular interest is also (2-nitroethyl)-benzene (**41**), a rare nitro-group containing natural product with cinnamon scent and known from the flower odor of several plants.[50] It can be regarded as an oxidized form of benzyl cyanide (**42**), present in some individuals and known as an anti-aphrodisiac pheromone of *Pieris* butterflies.[51]

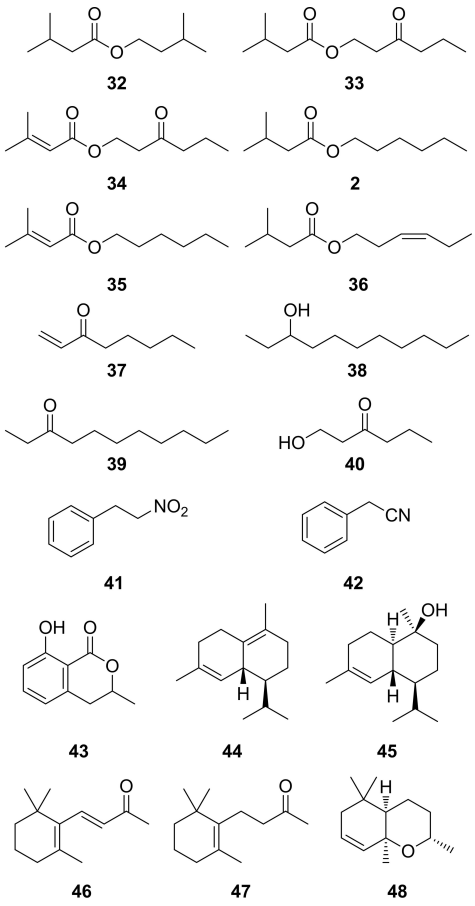


Figure 9. Volatile compounds form the CSG gland of *H. erato*.

Several other compound classes of lower importance include long chain alcohols such as 1-docosanol or tetracosen-1-ol, as well as respective diols, e.g. 1,3-docosanediol and 1,3-tetracosanediol. The latter compounds can be detected by their

Table 3. Gas chromatographic retention indices *I* of alcohol terpenes.

terpene	R–OH	*I*	R–OH	*I*
[2]	geraniol/nerol[a]	1260/1228	linalool[a] (**49**)	1101
[3]	(*E*,*E*)-farnesol (**4**)	1721	(*E*)-nerolidol (**52**)	1565
[4]	(*E*,*E*,*E*)-geranylgeraniol (**5**)	2190	(*E*,*E*)-geranlylinalool (**54**)	2026
[5]	(*E*,*E*,*E*,*E*)-geranylfarnesol (**58**)	2660	(*E*,*E*,*E*)-geranylnerolidol (**57**)	2486
[6]	(*E*,*E*,*E*,*E*,*E*)-farnesylfarnesol (**30**)	3132	(*E*,*E*,*E*,*E*)-farnesylnerolidol (**31**)	2966
[7]	(*E*,*E*,*E*,*E*,*E*,*E*)-geranylgeranylfarnesol (**62**)	3602	(*E*,*E*,*E*,*E*,*E*)-geranylgeranylnerolidol (**61**)	3432

[a] Literature values obtained from Ref. [38].

typical gas chromatographic behavior, reaction with silicone GC-phases, leading to cyclic dimethylsiloxane derivatives.[52] Other compounds, also regularly found in lepidopteran scent organs, were cholesterol, tocopherol, acting likely as stabilizer to prevent degradation of labile constituents due to reactive oxygen, and squalene. In contrast to the linear terpenes mentioned above, squalene shows a conventional [3+3]-terpene structure.

Typical constituents of the cuticle of butterflies, linear alkanes and alkenes, are also present in the CSG, e.g. heptacosane, but also highly methyl-branched compounds such as 7,11,15- and 9,13,17-trimethyl-, 5,9,13,17- and 7,11,15,19-tetramethyl-nonacosanes and -hentriacontanes. The structural characterization is shown exemplary in Figure S9 in the Supporting Information. In contrast to the linear alkanes, branched alkanes occur also on the wing area controls of the butterflies without scent scales, together with long chain aldehydes. This indicates that they are common cuticular components and not CSG specific. The same is true for 2,5-dialkyltetrahydrofurans such as 2-nonyl-5-octadecyltetrahydrofuran, commonly found on the cuticle of butterflies.[53]

About 550 compounds of the 775 detected occurred at least three times within the samples, of which 360 were identified. The unknowns were, to a large extent, esters with a known acyl part or hydrocarbons, usually occurring in small amounts. The Top 50 compounds found in highest mean concentration within the samples are listed in Table 1, whereas a full list of compounds can be found in the SI, Table S7. Usually, the esters **8l** and **31** are major components of the secretion. A single individual can have up to 350 components, thus showing a relatively high individual compositional variation of the CSG. In Table 4 the Top 50 volatile (Top 50 voc) compounds are shown, because only three volatiles occur in Table 1. The most important volatiles here are **1**, **39** and **38**, originating from very different biosynthetic pathways. The high number of components is unprecedented in the glandular chemistry of butterflies.

Table 4. The 50 volatile compounds (Top 50 voc) occurring in highest mean amounts in CSG extracts with I < 1730. The mean value were calculated excluding samples that did not contain the compound. *I*: gas chromatographic retention index; oc: number of occurrence in the 114 samples analyzed; µg: mean amount of the compound in CSG extracts containing them. A minimum of oc = 3 was required for listing. A full list of compounds can be found in the SI.

	Compound	*I*	oc	µg
1	3-Undecanone (**39**)	1283	99	5.65
2	(*E*)-β-Ocimene (**1**)	1043	115	3.10
3	1-Hexen-3-one	761	3	2.56
4	2-Nitroethylbenzene (**41**)	1281	101	0.66
5	1-Octen-3-one (**37**)	973	106	0.38
6	Mellein (**43**)	1524	67	0.28
7	(2*E*,6*E*)-α-Farnesene	1505	17	0.28
8	1-Decen-3-one	1174	10	0.27
9	Unknown B135_1	1389	112	0.24
10	Hexyl 3-methylbutyrate (**2**)	1238	59	0.22
11	Benzyl cyanide (**42**)	1126	99	0.22
12	(*E*)-Farnesene-9,10-epoxide	1621	8	0.21
13	3-Nonanone	1081	93	0.17
14	Dihydro-β-ionone (**47**)	1427	75	0.16
15	(*E*)-Nerolidol	1554	6	0.15
16	Unknown B98_1	1015	5	0.15
17	(*Z*)-3-Hexenyl 3-methylbutyrate (**36**)	1232	60	0.15
18	Undecen-3-one	1281	76	0.11
19	2-Phenylethyl 3-methyl-2-butenoate	1565	49	0.10
20	(*Z*)-3-Hexenyl tiglate	1306	66	0.10
21	Undecadien-3-one	1283	13	0.07
22	Unknown B95_3	1717	40	0.06
23	3-Undecanol (**38**)	1292	85	0.05
24	3-Oxohexyl 3-methylbutanoate (**33**)	1386	42	0.05
25	Germacrene-D	1467	36	0.05
26	Dihydroedulan-II (**48**)	1279	99	0.05
27	Unknown B59	1189	10	0.05
28	3-Decanone	1181	56	0.04
29	Unknown alkenol_1	1274	27	0.04
30	(*E*)-β-Ionone (**46**)	1474	43	0.04
31	Isogeranial	1176	28	0.04
32	T-Cadinol (**45**)	1628	72	0.03
33	Unknown B95_1	1426	19	0.03
34	Unknown B81_1	1488	10	0.03
35	Unknown alkene_1	943	51	0.03
36	3-Oxohexyl 3-methyl-2-butenoate (**34**)	1461	23	0.03
37	α-Terpinyl acetate	1338	35	0.03
38	Alloocimene	1133	4	0.03
39	3-Methylbutyl 3-methylbutanoate (**32**)	1103	23	0.03
40	Undecenyl acetate	1489	17	0.03
41	γ-Cadinene (**44**)	1503	34	0.03
42	α-Cadinol	1642	14	0.02
43	Methoxyphenyl-oxime	899	34	0.02
44	Dihydroedulan I	1286	12	0.02
45	Unknown 3-oxohexyl ester (1)	1398	11	0.02
46	Humulenepoxide II	1597	4	0.02
47	2-Cyclopentene-1-carboxylic-acid	1009	11	0.02
48	Caryophyllene oxide	1572	6	0.02
49	1-Hepten-3-one	880	24	0.02
50	(*E*)-2-*epi*-β-caryophyllene	1449	12	0.02

Biosynthesis

Previously it was shown in *H. melpomene* that pheromone gland composition in *Heliconius* is mostly independent from host plants, which are always *Passiflora* sp., and not influenced by adult pollen feeding.[16] Only sesquiterpenes, compounds like **44** or **45**, usually present in only low amounts in CSGs, were dependent on the larval host plant, as was the case for some aromatic androconial compounds. It seems therefore likely that most of the CSG compounds are biosynthesized by *H. erato* itself, different to e.g. danaines and ithomiines which rely on pyrrolizidine alkaloid uptake.[54,55] A key transformation is obviously the ligation of acid precursors and alcohols[34] to arrive at the large variety of esters. The acids themselves originate either from the fatty acid pathway or form the terpene pathway.

The terpene pathway in *H. melpomene* has recently been investigated by us and the (*E*)-β-ocimene (**1**) synthase HmelOS has been functionally characterized.[16] HmelOS is part of a family of seven geranylgeranyl pyrophosphate synthases (GGPPS) present in the genome. The ortholog HcydOS in the sister species *H. cydno* cannot produce **1** due to sequence differences. Homologues to HmelOS and other *H. melpomene* GGPPS also exist in *H. erato* (Figure S38 in the SI). The unique activity of HmelOS to produce the conjugated dienyl head group of **1** seems to be active also in the *H. erato* enzymes leading to various analogs of this type, although the actual

enzyme is not known. Furthermore, the elongation capability is active within this family, leading to the characteristic longer terpene hydrocarbons. Terpene elongation can of course also take place by other enzymes, e.g. decaprenyl synthases also known from *Heliconius*.[16]

The biosynthesis of the major CSG terpenes can be explained as shown in Figure 10. A pyrophosphate precursor of different chain length can be transformed into terpene hydrocarbons of the α- or β-farnesene type. In the case of geranyl pyrophosphate (GPP), only the α-type **1** is formed,[16] confusingly named (*E*)-β-ocimene in contrast to both the α- and β-type producing higher pyrophosphates. Alternatively, hydrolysis can lead to tertiary alcohols of the nerolidol type,[16] whereas a phosphatase releases the primary alcohols, usually present in higher amounts than the tertiary alcohols. In case of farnesyl pyrophosphate (FPP), the respective alcohol (*E,E*)-farnesol (**4**) serves as a precursor for acid **8a**, requiring additional oxidation and hydrogenation. The same transformations are observed for higher pyrophosphates.

The biosynthesis of other CSG compounds is not known in *Heliconius*, but their structures allow assignments to established biosynthetic pathways, described for example by Morgan,[56] taken into account known pathways of other organisms. The complex chemical bouquet of the *H. erato* CSG can be understood by analyzing their biosynthetic pathways and some key transformations. Only a few specific transformations are responsible for the observed high structural variation (Figure 11).

The terpene and the fatty acid biosynthesis pathways deliver building blocks for the key ester formation that also uses short alcohols recruited from different pathways. Although the major compounds seem to originate from butterfly biosynthesis, contributions from food or microorganisms cannot be ruled out, as discussed above.

The large number of compounds can be divided into various compound classes according to their biosynthetic

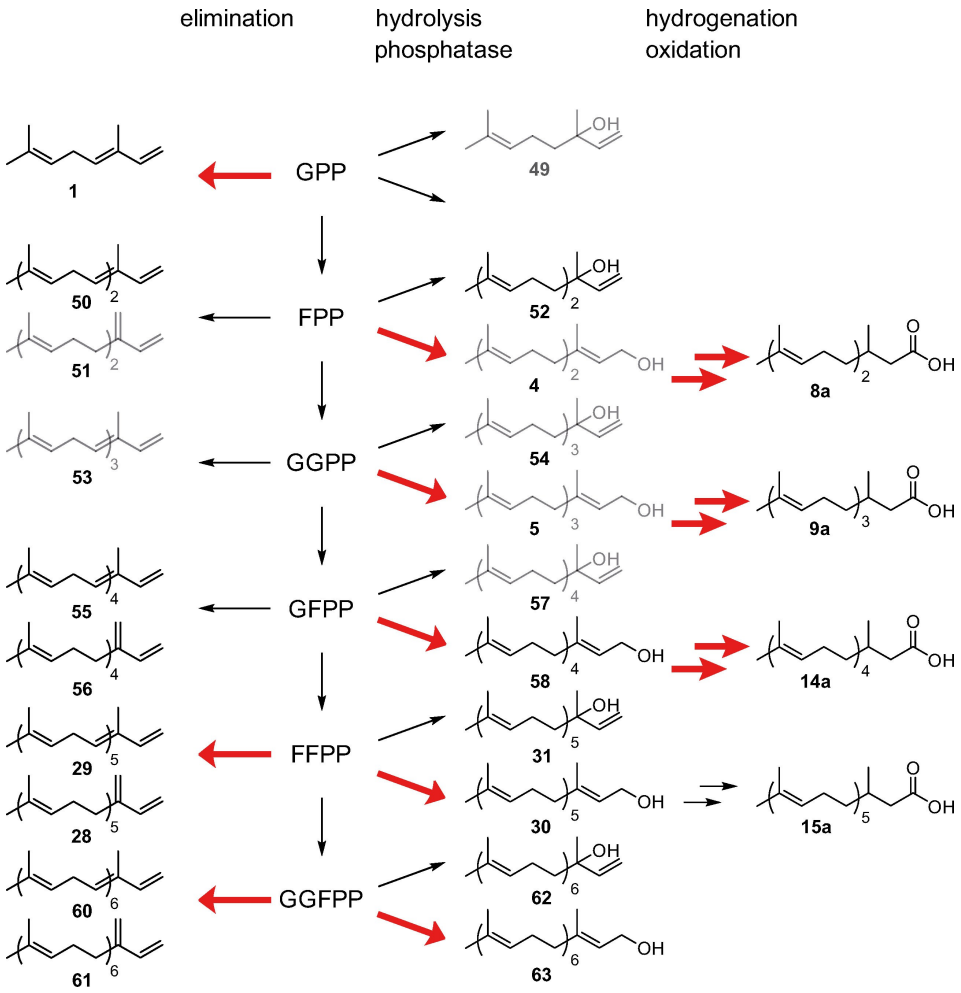


Figure 10. Biosynthesis of terpenes in *H. erato*. Major pathways are indicated by red arrows and minor components of the CSG secretion are shown in grey. Typical terpene elongation takes place from geranyl pyrophosphate (GPP) via farnesyl pyrophosphate (FPP) to geranylgeranyl pyrophosphate (GGPP). Further elongation is performed by unusual enzyme activity leading to geranylfarnesyl pyrophosphate (GFPP), farnesylfarnesyl pyrophosphate (FFPP), and geranylgeranylfarnesyl pyrophosphate (GGFPP). Elimination leads to α- and β-terpene hydrocarbons, whereas hydrolysis delivers alcohols of the nerolidol (**54**) type. A phosphatase furnishes terminal alcohols that will be in turn further modified by hydrogenation and oxidation to form dihydroterpene acids, precursors of the CSG esters.

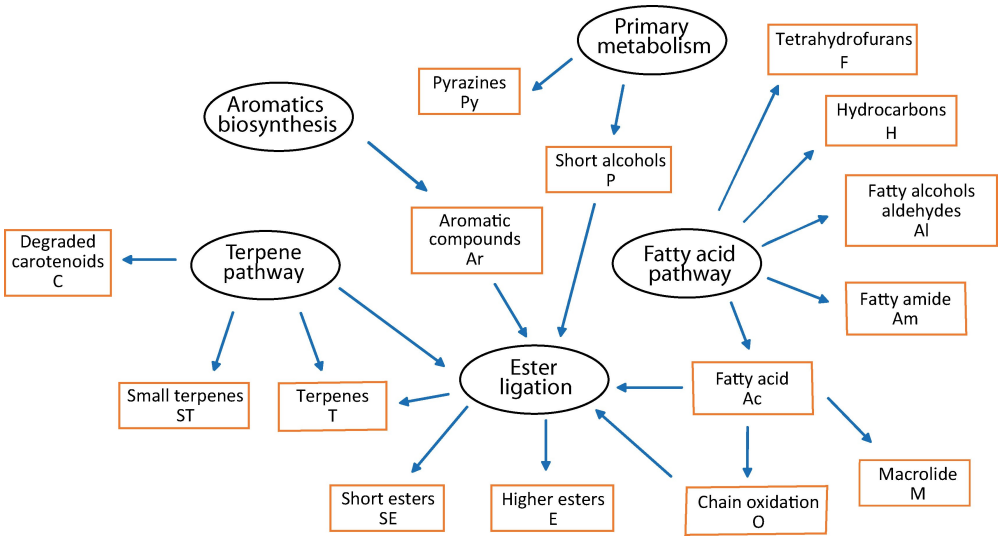


Figure 11. Definition of compound classes (brown boxes) based on biosynthetic pathways of CSG compounds of *H. erato*.

pathway as shown in Figure 11. By this approach, we defined 15 different compound classes. Fatty acids (Ac) serve as precursors for aldehydes and alcohols via reduction (Al), leading for example to 1-tetracosanol. Linear and branched hydrocarbons (H) such as 7,11,15-trimethylnonacosane are obtained in insects by fatty acid biosynthesis followed by loss of the terminal carboxylate group.[57] The biosynthesis of 2,5-dialkyltetrahydrofurans (F) is unknown, but their chain length[53] indicates fusion of two common fatty acids to arrive at their carbon skeleton. Amides (Am) such as eicosanoic acid amide can be formed by amidation of acids, whereas diols such as 1,3-docosanediol are assigned to the Al class, likely formed by reduction of 3-hydroxy acids. Chain oxidation of fatty acids can take place at two different sites. Whereas mostly near terminal oxidation followed by ring closure leads to macrolides (M) with the same carbon number as the parent acid,[58] internal oxidation by lipoxygenases and chain cleavage eventually leads to short chain cleavage products (O) such as the leaf alcohol (*Z*)-3-hexenol.[56] An ester ligase now combines acids and these and various alcohols originating from the primary metabolism such as ethanol (P), or from the terpene pathway (T), e.g. 3-methyl-3-butenol. Aromatic alcohols (Ar) such as 2-phenylethanol are also found free or bound as esters. Although aromatic compounds can be formed by various pathways, it will not be divided here because pathway attribution is not always evident form the structure itself and include likely microorganism metabolites such as mellein. Minor compound classes comprise degradation or side products of carotenoid biosynthesis (C) such as dihydroionone (**47**) or defensive pyrazines (Py).[59,60] The terpene pathway (T) has been discussed above. To account for the largely different volatility within esters, a second esters class, short esters such as hexyl 3-methybutanoate (**2**, SE), was established. Similarly, small terpenes (ST) include volatile terpenes such as **1**.

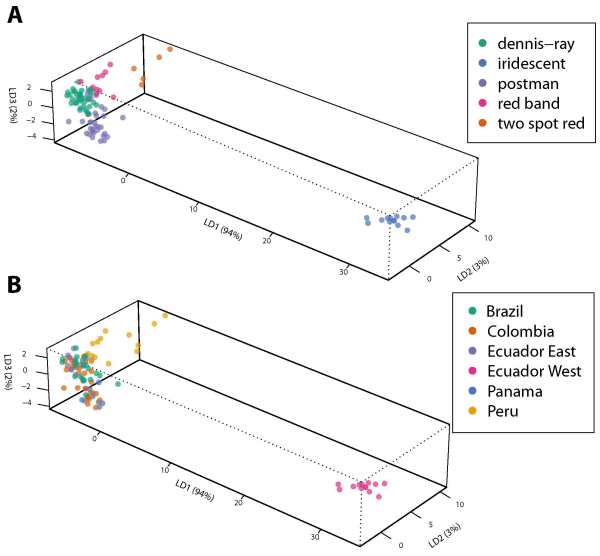


Figure 12. DAPC analysis of different mimicry rings of all compounds occurring in at least three samples, normalized within individuals. A: DAPC analysis based on mimicry. B: Results of A color-coded according to the origin of the samples.

Data analysis

The results of the analysis of the individual CSG composition were then used to test whether mimicry group membership is related to CSG composition. A discriminant analysis of principal components (DAPC) was performed in R using the package adegenet.[61,62] All components occurring in at least three samples were used, whereas hybrids were excluded. The results are shown in Figure 12 and Figures S39 and S40 in the SI.

Our earlier analysis revealed geographic variation in the CSG contents between the samples from West Ecuador and the other locations.[13] The former butterflies constitute the iridescent mimicry group (Figure 2). The iridescent group is also

clearly differentiated from all four other mimicry groups included in our analysis, dennis-ray, postman, red band, and two spot red (Figure S39). Nevertheless, these four mimicry groups are also separated, but some overlap occurs between postman and dennis-ray. When the results are used as predictive model to assign samples to the different mimicry patterns, the outcome is less convincing (Figure S41). Again, iridescent as well as red band groups showed good agreement, whereas assignment to the other groups were weak.

The separation is better when normalization within the samples was performed (Figure S40A the SI), effectively comparing relative proportions of compounds of samples. The normalization removes variation in the absolute amount of the gland compounds, which is often observed in samples from the wild where the life history and physiological state of the butterflies cannot be controlled. Therefore, we used normalized data in all following DAPC analyses. The model now allows good predictions of the mimicry group for individuals in most cases, although differentiation between postman and dennis-ray is not always possible (see SI, Figure S42). The observed separations might depend also on geographic variation and might actually not be connected to mimicry groups. To test this, we colored the DAPC results according to geographic origin which reveals some insights (Figure 9B). Whereas the iridescent group was collected only in West Ecuador, dennis-ray and postman were sampled from different locations. The results show that the separation is not likely to be connected to the geographic origin. Furthermore, red band and two spot red individuals originate almost completely from Peru, but still show different clusters in the DAPC analysis.

The CSG content consists largely of compounds of low or no volatility (Figure 3), whereas volatile compounds are less concentrated. Nevertheless, biological activity has been described, especially for volatile compounds.[5,6,9] Therefore, we asked whether there might be differences between mimicry groups when comparing compounds within different volatility ranges only. The retention index *I* of an apolar GC column such as HP-5 can be used as a rough approximation for volatility[63] and was used here to define three groups of compounds. Volatile compounds included all compounds with $I < 1750$, including methyl 2,3-dihydrofarnesoate (**8b**), but excluding dihydrofarnesoic acid (**8a**, *I* 1768), because acids have usually lower volatility as estimated from their *I* value. A group of low volatility followed with a cutoff at *I* 2,500, as used in earlier publications.[8,10,15] And finally a non-volatile region with $I > 2{,}500$ was defined, extending up to about *I* 4,000.

The results of the DAPC of these groups are shown in Figure S40. In all classes the iridescent group is separated from all other, most effectively in the non-volatile group. Two spot red is also separated in all groups. Dennis-ray and postman separate within the volatiles but are not well distinguishable in the two other groups. Red band can be effectively predicted in both the volatile and non-volatile groups (see outcome of model predictions in Figures S41–43 in the SI). In summary, the volatile compounds in particular are effective at separating all mimics, with few exceptions, and have good model predictability (Figure S43).

We were also interested in the compounds responsible for the separation of the different mimetic groups. Heatmaps of the compounds of Table 1 (Top 50) and of the 50 most important volatile compounds (Top 50 voc, Table 4) are shown in Figures S46 and S47. The iridescent mimics are clearly separated from the others by high concentrations of a range of 2-phenylethyl esters. In addition, the higher terpene alcohols such as **30** and the respective hydrocarbons **28** and **29** are absent. In the volatile region the important ketone 3-undecanone (**39**) is missing, whereas instead benzyl cyanide (**42**) occurs in higher concentrations. A short phenylethyl ester, 2-phenylethyl 3-methyl-2-butenoate is missing in the other mimicry groups. Nevertheless, the differentiation of the other mimicry groups is not easily detectable in the heatmaps.

Some insight into the discriminating compounds can be derived from their loadings contributing to the linear discriminants (Tables S1–S5). The iridescent mimics are separated due to 2-phenylethyl tetradecanoate and dodecanoate as well as 3-oxohexyl esters in linear discriminant 1 (LD1) including all compounds by concentration, many of them being among the Top 50 (Table S1). LD2 is largely dependent on hydrocarbons that do not occur within the Top 50. The situation changes if normalized data are used (Table S2). Both benzyl cyanide (**42**) and 3-oxohexyl tetradecanoate (**18l**) contribute most, whereas now many minor compounds with low average concentrations contribute to all dimensions. The hydrocarbons become less important, whereas volatile compounds contribute more. In the volatile group (Table S3) many minor components contribute most, with some Top 50 voc compounds such as 3-methylbutyl 3-methylbutanoate (**32**) in LD1. Of major importance in LD2 and LD3 are some carotenoids such as β-cyclocitral or 7,8-dihydro-β-ionol. Loadings in the low-volatile group (Table S4) indicate important contributions again due to **18l**, and 3-methylbutyl (*E*)-2,3-dihydrofarnesoate (**8h**), the latter also important in LD3. Most separation in LD2 is due to isoprenyl (*E*)-2,3-dihydrofarnesoate (**8f**), whereas **57** additionally is important in LD3. In the non-volatile group (Table S5) important separating compounds are 1,3-tetracosanediol in LD1, an unknown (*E*)-2,3-dihydrofarnesylfarnesoate ester in LD2, and the hydrocarbon 7,11,15-trimethylnonacosane in LD3.

The heatmap revealed clear separation between the iridescent and all other mimics. Therefore, we analyzed the dataset again asking whether there is differentiation between iridescent and all other mimics taken together. The loadings revealed now similar results as the heatmap, showing that 2-phenylethyl esters **18p**, **19p**, **23p**, and **8p** are responsible for more of 50% of the separation (Table S6), followed in importance by **39**.

The results of these DAPC analyses indicate that the mimicry groups are indeed chemically differentiated according to their CSG gland contents. Because most mimicry groups are collected from at least two geographic regions, a masking regional influence on mimicry seems not to be dominant. As found in the earlier geographic analyses,[13] the iridescent group is quite distant from the other mimicry groups, whereas postman and dennis-ray are very close. Whether these differences are used by the butterflies for example for species recognition, and

whether the composition is influenced by other species in the respective mimicry group to secure coherence in the group, e.g. by a general mimicry odor, is not clear. Nevertheless, the specific chemistry used by *H. erato*, namely the specific linear terpenes, the unusual alcohols such as 3-oxohexanol and rare compounds like nitroethylbenzene hint at an elaborate chemical signaling involving the CSG secretion beyond known functions.

The identification of many of the unknown compounds in CSGs of *H. erato* allowed assignment of these compounds in the other species investigated by us earlier, *H. sapho, eleuchia, elevatus, melpomene, timareta* and *cydno*.[13] The thus updated compound list of the previous publication (see SI, Table S6) was used to assign each compound to one of the biosynthetic classes defined above. Then summation of the average amount produced of all compounds of a certain class was performed (see SI, Table S6). In cases of esters formed from two different classes the amount was divided 2:1 in favor of the larger component. The resulting proportions of the summed average amounts of the different biosynthetic classes are shown in Figure 13, revealing similarities and differences in the usage of the pathways. In some cases closely related species tap similar compound classes, as is the case for *H. sapho/eleuchia* and slightly less evident, *H. timareta/cydno*. *Heliconius elevatus/melpomene*, are also quite similar, although they are less close to each other. These pairs differ between each other and to *H. erato*. As described, terpenes dominate in the latter, with **8l**, **28**, and **30** as major components. The *H. elevatus/melpomene* pair also uses preferentially terpenes, but almost exclusively **1** (ST), also present in reduced amount in *H. erato*. Hydrocarbons are the second largest group in this pair, with heneicosane in both species, but *H. elevatus* additionally produces larger amounts of heneicosene and tricosene. The dominating aromatic compound in *H. sapho/eleuchia* is benzyl salicylate (Ar), whereas other important compounds are **47** (C) in both species and hexyl esters of fatty acids (E, O) only in *H. eleuchia*. *Heliconius timareta* and *H. cydno* are not so similar, with 9,11-octadecadien-13-olide (M) as major component in the latter.[11] *Heliconius timareta* is the only other species with a large M group, but with 9-octadecen-13-olide as the most important compound of this class. Additionally, many fatty acid esters with short alcohols are present, e.g. isopropyl and butyl oleate (P, Ac). All of the compounds mentioned here except the macrolides are also present in *H. erato*, although in small amounts only. This indicates the genetic similarity of the species. Probably different regulation leads to the species-specific mixtures. Nevertheless, our in depth analyses of HmelOS indicated functional differences between similar genes in *H. melpomene* and *H. cydno*.[16] These data show the remarkable biosynthetic plasticity of *Heliconius* CSG chemistry.

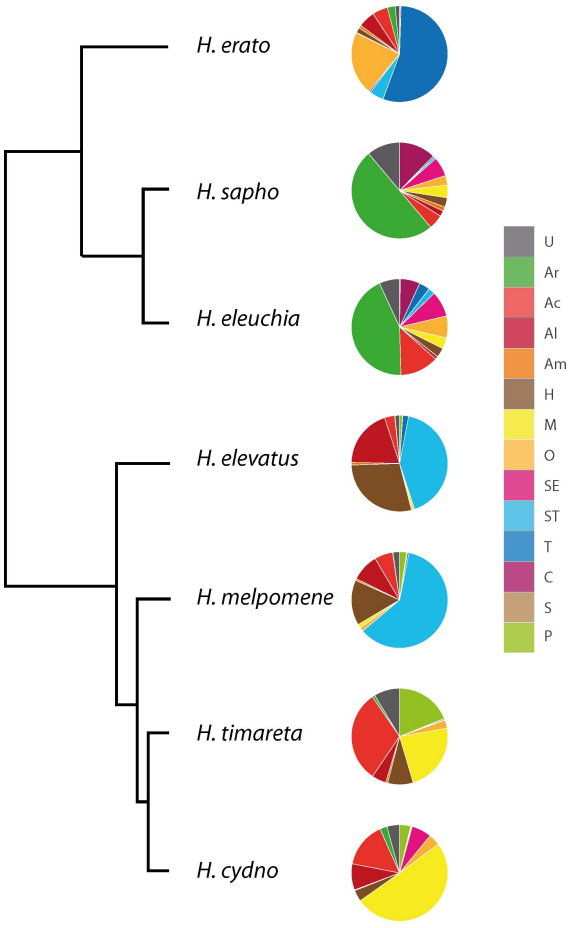


Figure 13. Distribution of biosynthetic classes within *Heliconius sapho*, *H. eleuchia*, *H. elevatus*, *H. melpomene*, *H. timareta*, *H. cydno*, and *H. erato*. Data obtained by summation of average amounts of compounds attributed to each class taken from Darragh et al.[13] and phylogeny after Kozak et al.[66] Compound class definition see text.

Given the different habitats and ecological niches used by the heliconiines it seems unlikely that major CSG compounds or their direct precursors are directly taken up pharmacophagously or with food, as is the case in other butterflies such as ithomiines and danaines.[55,64] The high compound diversity and complex chemistry of the CSG hints towards more semiochemical functions of the secretion beyond the known roles in anti-aphrodisiac signaling and physical evaporation rate modification. Such traits might include species recognition, defense,[13,60,65] territorial signals, physiological changes in the female, or even antimicrobial activity. These roles can only be tested experimentally. As has been the case repeatedly in insects, it has to be mentioned that minor components in a secretion may have strong semiochemiocal activity, mandating experiments even with minor components.

Conclusion

By combining different microanalytical techniques with synthesis we identified numerous new natural products produced in the CSG of male *H. erato*. The terpene pathway is extensively used to produce a wide array of more than 500 compounds. Key aspects are the extension of known linear terpene biosynthesis up to [7]-terpenes, formation of dihydroterpene compounds and ester formation. Ester alcohols include unique compounds such as 3-oxohexanol. The metabolomic analyses of the data revealed a high differentiation of the iridescent

mimicry group from the other, also differentiated, mimicry groups, especially within the volatile compounds. Because many of the compounds occur in other species as well, phylogenetic preference for different compound groups defined by biosynthetic transformations was shown. Our analysis revealed that the thorough analyses of metabolite composition serves as a basis for the understanding of the composition of chemical signals in complex mimicry groups.

Experimental Section

Samples

The samples investigated here originate from *H. erato* butterflies collected in 2016 and 2017 in the field as described.[13] They were collected from various locations in Brazil, Colombia, Ecuador, Panama, and Peru. Additionally, three samples of *H. erato lativitta* collected in Tena, Ecuador in 2019 were analyzed. Butterflies were collected with relevant local collection permits. Clasper scent glands were extracted for 1 h with 200 µl CH_2Cl_2 containing 200 ng 2-tetradecyl acetate as internal standard directly after dissection in the field. The samples were stored at low temperatures throughout transport and until analyses. Full details of sample collection and treatment have been described earlier.[13]

Analysis

GC/MS data from the initial analyses[13] were used. Extracts were analyzed by GC/MS using an Agilent 5977 mass-selective detector connected to an Agilent 7890B gas chromatograph and equipped with an Agilent ALS 7693 autosampler. HP-5MS fused silica capillary columns (Agilent, 30 m × 0.25 mm, 0.25 µm) were used. Injection was performed in splitless mode (250 °C injector temperature) with helium as the carrier gas (constant flow of 1.2 ml/min).[13] The additional samples were analyzed using an Agilent model 5975 mass-selective detector connected to an Agilent 7890A gas chromatograph and equipped with an Agilent ALS 7683B autosampler equipped with an identical column. The temperature program started in all analyses at 50 °C, was held for 5 min, and then rose at a rate of 5 °C/min to 320 °C, before being held at 320 °C for 5 min.[13] GC/DD-IR spectra were measured with a Dani Instruments DiscovIR IR-detector coupled to an Agilent Technologies 7890B gas chromatograph equipped with an identical capillary column as used for GC/MS analyses. Components were identified by comparison of mass spectra and gas chromatographic retention index of commercial and in-house databases, the latter obtained from authentic reference samples. Enantiomer determinations were performed using a column with a Hydrodex-6-TBDM (heptakis-(2,3-di-*O*-methyl-6-*O*-*t*-butyldimethylsilyl)-β-cyclodextrin) phase.

Statistical analyses

A data file containing relative concentrations for each compound in every analysis against the internal standard was prepared. The data were curated by excluding artifacts and contaminations such as phthalates, adipates, or silicon compounds. Zero values were substituted with 0.0001, half the amount of the smallest component.[67] Then a discriminant analysis of principal components (DAPC) testing against mimicry groups in R was performed, using the package adegenet.[61,62] The number of retained principal components was optimized using the a-score and cross-validation to avoid overfitting.[68] Finally, membership probability was tested for each sample, as shown in Figures S41–45. Heatmaps were obtained with the package pheatmap,[69] and pie charts with standard R. The R code used can be found in the Supporting Information.

General synthetic methods

Full experimental procedures can be found in the Supporting Information.

Supporting information

The Supporting Information contains mass spectra, IR-spectra, synthetic procedures, total ion chromatograms, statistical analyses, R code, A-scores, heatmaps. Table S7 of identified compounds. Tables S8 and S9 of compounds reassigned from Darragh et al.[13]

Acknowledgements

We thank the Deutsche Forschungsgemeinschaft for funding in the project "Mimicry", Schu984/12-1. Open Access funding enabled and organized by Projekt DEAL.

Conflict of Interest

The authors declare no conflict of interest.

Keywords: biosynthesis · GC/IR · pheromones · terpenoids · volatile esters

[1] a) H. W. Bates, *Trans. Linn. Soc. London* **1862**, *23*, 495–566; b) C. D. Jiggins, *The Ecology & Evolution of Heliconius butterflies*, Oxford University Press, Oxford, **2017**.
[2] a) A. Nahrstedt, R. H. Davis, *Comp. Biochem. Physiol. Part B* **1985**, *82*, 745–749; b) H. S. Engler-Chaouat, L. E. Gilbert, *J. Chem. Ecol.* **2007**, *33*, 25–42; c) É. C. Pinheiro de Castro, R. Demirtas, A. Orteu, C. E. Olsen, M. S. Motawie, M. Zikan Cardoso, M. Zagrobelny, S. Bak, *Insect Biochem. Mol. Biol.* **2020**, *116*, 103259.
[3] T. N. Sherratt, *Naturwissenschaften* **2008**, *95*, 681–695.
[4] a) T. Ando, S. Inomata, M. Yamamoto in *The Chemistry of Pheromones and Other Semiochemicals I, Topics in Current Chemistry, Vol. 239*, (Ed.: S. Schulz), Springer, Heidelberg, **2004**, pp. 51–96; b) R. I. Vane-Wright, M. Boppré, *Phil. Trans. R. Soc. B* **1993**, 197–205.
[5] L. E. Gilbert, *Science* **1976**, *193*, 419–420.
[6] S. Schulz, C. Estrada, S. Yildizhan, M. Boppré, L. E. Gilbert, *J. Chem. Ecol.* **2008**, *34*, 82–93.
[7] K. Darragh, S. Vanjari, F. Mann, M. F. Gonzalez-Rojas, C. R. Morrison, C. Salazar, C. Pardo-Diaz, R. M. Merrill, W. O. McMillan, S. Schulz, C. D. Jiggins, *PeerJ* **2017**, *5*, e3953.
[8] K. J. R. P. Byers, K. Darragh, J. Musgrove, D. A. Almeida, S. F. Garza, I. A. Warren, P. M. Rastas, M. Kučka, Y. F. Chan, R. M. Merrill, S. Schulz, W. O. McMillan, C. D. Jiggins, *Evolution* **2020**, *74*, 349–364.
[9] C. Estrada, *Sexual Behavior, Intraspecific Signaling and the Evolution of Mimicry Among Closely Related Species*, PhD Thesis, Austin, **2009**.
[10] F. Mann, S. Vanjari, N. Rosser, S. Mann, K. K. Dasmahapatra, C. Corbin, M. Linares, C. Pardo-Diaz, C. Salazar, C. Jiggins, S. Schulz, *J. Chem. Ecol.* **2017**, *43*, 843–857.
[11] S. Schulz, S. Yildizhan, K. Stritzke, C. Estrada, L. E. Gilbert, *Org. Biomol. Chem.* **2007**, *5*, 3434–3441.

[12] C. Estrada, S. Schulz, S. Yildizhan, L. E. Gilbert, *Evolution* **2011**, *65*, 2843–2854.
[13] K. Darragh, G. Montejo-Kovacevich, K. M. Kozak, C. R. Morrison, C. M. E. Figueiredo, J. S. Ready, C. Salazar, M. Linares, K. J. R. P. Byers, R. M. Merrill, W. O. McMillan, S. Schulz, C. D. Jiggins, *Ecol. and Evol.* **2020**, *10*, 3895–3918.
[14] a) M. F. González-Rojas, K. Darragh, J. Robles, M. Linares, S. Schulz, W. O. McMillan, C. D. Jiggins, C. Pardo-Diaz, C. Salazar, *Proc. R. Soc. London Ser. B* **2020**, *287*, 20200587; b) C. Mérot, B. Frérot, E. Leppik, M. Joron, *Evolution* **2015**, *69*, 2891–2904.
[15] K. Darragh, K. J. R. P. Byers, R. M. Merrill, W. O. McMillan, S. Schulz, C. D. Jiggins, *Ecol. Entomol.* **2019**, *44*, 397–405.
[16] K. Darragh, A. Orteu, D. Black, K. J. R. P. Byers, D. Szczerbowski, I. A. Warren, P. Rastas, A. Pinharanda, J. W. Davey, S. Fernanda Garza, D. Abondano Almeida, R. M. Merrill, W. O. McMillan, S. Schulz, C. D. Jiggins, *PLoS Biol.* **2021**, *19*, e3001022.
[17] R. J. Challis, S. Kumar, K. K. Dasmahapatra, C. D. Jiggins, M. Blaxter, *bioRxiv* **2016**; https://doi.org/10.1101/056994.
[18] S. Yildizhan, S. Schulz, *Synlett* **2011**, 2831–2833.
[19] A. Wang, B. Wüstenberg, A. Pfaltz, *Angew. Chem. Int. Ed.* **2008**, *47*, 2298–2300; *Angew. Chem.* **2008**, *120*, 2330–2332.
[20] M. Menke, K. Melnik, P. S. Peram, I. Starnberger, W. Hödl, M. Vences, S. Schulz, *Eur. J. Org. Chem.* **2018**, 2651–2656.
[21] R. Ryhage, E. Stenhagen, *Ark. Kemi* **1960**, *15*, 291–304.
[22] D. M. Vidal, C. F. Fávaro, M. M. Guimarães, P. H. G. Zarbin, *J. Braz. Chem. Soc.* **2016**, *27*, 1506–1511.
[23] a) K.-D. Müller, H. Husmann, H. P. Nalik, G. Schomburg, *Chromatographia* **1990**, *30*, 245–248; b) S. P. Chinta, S. Goller, G. Uhl, S. Schulz, *Chem. Biodiversity* **2016**, *13*, 1202–1220.
[24] a) H. Y. Ho, J. G. Millar, *J. Chem. Ecol.* **2001**, *27*, 1177–1201; b) H. Y. Ho, J. G. Millar, *J. Chem. Ecol.* **2001**, *27*, 2067–2095.
[25] T. Schmitt, E. Strohm, G. Herzner, C. Bicchi, G. Krammer, F. Heckel, P. Schreier, *J. Chem. Ecol.* **2003**, *29*, 2469–2479.
[26] J. C. Snyder, Z. Guo, R. Thacker, J. P. Goodman, J. S. Pyrek, *J. Chem. Ecol.* **1993**, *19*, 2981–2997.
[27] J. Wei, Le Kang, *Plant Signaling Behav.* **2011**, *6*, 369–371.
[28] C. Zdero, F. Bohlmann, R. M. King, H. Robinson, *Phytochemistry* **1986**, *25*, 2841–2855.
[29] T. J. Schmidt, S. Rzeppa, M. Kaiser, R. Brun, *Phytochem. Lett.* **2012**, *5*, 632–638.
[30] K. Preece, R. Glávits, J. R. Foster, T. Murbach, J. R. Endres, G. Hirka, A. Vértesi, E. Béres, I. P. Szakonyiné, *Regul. Toxicol. Pharmacol.* **2021**, *124*, 104975.
[31] J. E. Bello, P. Stamm, H. P. Leinaas, S. Schulz, *Eur. J. Org. Chem.* **2019**, 2158–2162.
[32] P. Scribe, J. Guezennec, J. Dagaut, C. Pepe, A. Saliot, *Anal. Chem.* **1988**, *60*, 928–931.
[33] F. Mann, D. Szczerbowski, L. de Silva, M. McClure, M. Elias, S. Schulz, *Beilstein J. Org. Chem.* **2020**, *16*, 2776–2787.
[34] H.-L. Wang, O. Brattström, P. M. Brakefield, W. Francke, C. Löfstedt, *J. Chem. Ecol.* **2014**, *40*, 549–559.
[35] J. W. Grula, J. D. McChesney, O. R. Taylor, *J. Chem. Ecol.* **1980**, *6*, 241–256.
[36] S. Yildizhan, J. van Loon, A. Sramkova, M. Ayasse, C. Arsene, C. ten Broeke, S. Schulz, *ChemBioChem* **2009**, *10*, 1666–1677.
[37] P. A. A. Klusener, L. Tip, L. Brandsma, *Tetrahedron* **1991**, *47*, 2041–2064.
[38] P. Linstrom, "NIST Chemistry WebBook", can be found under http://webbook.nist.gov/chemistry/name-ser.html#Top, **2018**.
[39] T. Sato, H. Yamaga, S. Kashima, Y. Murata, T. Shinada, C. Nakano, T. Hoshino, *ChemBioChem* **2013**, *14*, 822–825.
[40] T. Ríos, S. C. Pérez, *J. Chem. Soc. D* **1969**, 214–215.
[41] S. Nozoe, M. Morisaki, K. Fukushima, S. Okuda, *Tetrahedron Lett.* **1968**, *9*, 4457–4458.
[42] a) T. Akihisa, K. Koike, Y. Kimura, N. Sashida, T. Matsumoto, M. Ukiya, T. Nikaido, *Lipids* **1999**, *34*, 1151–1157; b) W. J. Baas, *Phytochemistry* **1985**, *24*, 1875–1889; c) B. Singh, P. K. Agrawal, R. S. Thakur, *Phytochemistry* **1989**, *28*, 1980–1981; d) C. M. Feitosa, M. Z. B. Bezerra, A. M. d. G. L. Citó, J. S. d. Costa Júnior, J. A. D. Lopes, J. M. Moita Neto, *Quim. Nova* **2007**, *30*.
[43] J. C. Moreno, J. Mi, Y. Alagoz, S. Al-Babili, *Plant J.* **2021**, *105*, 351–375.
[44] W. Francke, S. Schulz, V. Sinnwell, W. A. König, Y. Roisin, *Liebigs Ann. Chem.* **1989**, 1195–1201.
[45] G. D. Prestwich, F. D. Whitfield, G. Stanley, *Tetrahedron* **1976**, *32*, 2945–2948.
[46] R. Nishida, T. C. Baker, W. L. Roelofs, *J. Chem. Ecol.* **1982**, *8*, 947–959.
[47] G. Kunesch, P. Zagatti, A. Pouvreau, R. Cassini, *Z. Naturforsch.* **1987**, *42c*, 657–659.
[48] F. Kern, R. W. Klein, E. Janssen, H.-J. Bestmann, A. B. Attygalle, D. Schafer, U. Maschwitz, *J. Chem. Ecol.* **1997**, *23*, 779.
[49] P. Reveglia, M. Masi, A. Evidente, *Biomol. Eng.* **2020**, *10*.
[50] J. L. Giongo, R. A. Vaucher, A. S. Da Silva, C. B. Oliveira, C. B. de Mattos, M. D. Baldissera, M. R. Sagrillo, S. G. Monteiro, D. L. Custódio, M. Souza de Matos, P. T. Sampaio, H. F. Teixeira, L. S. Koester, V. F. da Veiga Junior, *Microb. Pathog.* **2017**, *103*, 13–18.
[51] J. Andersson, A. K. Borg-Karlson, C. Wiklund, *J. Chem. Ecol.* **2003**, *29*, 1489–1499.
[52] S. Schulz, *Lipids* **2001**, *36*, 637–647.
[53] S. Schulz, G. Beccaloni, R. Nishida, Y. R. Roisin, I. Vane-Wright, J. N. McNeil, *Z. Naturforsch.* **1998**, *53c*, 107–116.
[54] a) J. A. Edgar, C. C. J. Culvenor, T. E. Pliske, *J. Chem. Ecol.* **1976**, *2*, 263–270; b) S. Schulz, W. Francke, M. Boppré, T. Eisner, J. Meinwald, *Proc. Natl. Acad. Sci. USA* **1993**, *90*, 6834–6838; c) T. E. Pliske, T. Eisner, *Science* **1969**, *164*, 1170–1172.
[55] S. Schulz, G. Beccaloni, K. S. Brown. Jr., M. Boppré, A. V. L. Freitas, P. Ockenfels, J. R. Trigo, *Biochem. Syst. Ecol.* **2004**, *32*, 699–713.
[56] E. D. Morgan, *Biosynthesis in insects*, Royal Society of Chemistry, Cambridge, **2010**.
[57] G. J. Blomquist in *Insect Hydrocarbons: Biology, Biochemistry, and Chemical Ecology* (Eds.: G. J. Blomquist, A. G. Bagnères), Cambridge University Press, Cambridge, **2010**, pp. 53–74.
[58] S. Schulz, S. Hötling, *Nat. Prod. Rep.* **2015**, *32*, 1042–1066.
[59] J. S. Dickschat, S. Wickel, C. J. Bolten, T. Nawrath, S. Schulz, C. Wittmann, *Eur. J. Org. Chem.* **2010**, 2687–2695.
[60] M. Rothschild, B. P. Moore, W. V. Brown, *Biol. J. Linn. Soc.* **1984**, *23*, 375–380.
[61] T. Jombart, *Bioinformatics* **2008**, *24*, 1403–1405.
[62] T. Jombart, I. Ahmed, *Bioinformatics* **2011**, *27*, 3070–3071.
[63] J. P. Durand, J. J. Beboulene, A. Ducrozet, A. Bre, S. Carbonneaux, *Oil Gas Sci. Technol.* **1999**, *54*, 431–438.
[64] a) G. W. Beccaloni, *Trop. Lepid.* **1997**, *8*, 103–124; b) J. A. Edgar, *J. Zool. Soc. (London)* **1982**, *196*, 385–399; c) J. Meinwald, Y. C. Meinwald, *J. Am. Chem. Soc.* **1966**, *88*, 1305–1310; d) T. Eisner, J. Meinwald, *Proc. Natl. Acad. Sci. USA* **1995**, *92*, 50–55.
[65] B. Rojas, E. Burdfield-Steel, H. Pakkanen, K. Suisto, M. Maczka, S. Schulz, J. Mappes, *Proc. R. Soc. London Ser. B* **2017**, *284*.
[66] K. M. Kozak, N. Wahlberg, A. F. E. Neild, K. K. Dasmahapatra, J. Mallet, C. D. Jiggins, *Syst. Biol.* **2015**, *64*, 505–524.
[67] S. C. Grace, D. A. Hudson in *Metabolomics: Fundamentals and Applications* (Ed.: J. K. Prasain), IntechOpen, London, **2016**, pp. 67–94.
[68] T. Jombart, C. Collins, "A tutorial for discriminant analysis of principal components (DAPC) using adegenet 2.0.0", can be found under http://adegenet.r-forge.r-project.org/files/tutorial-dapc.pdf, **2015**.
[69] R. Kolde, "Pheatmap: pretty heatmaps", can be found under https://scholar.google.com/citations?user=iyhbhfmaaaaj&hl=en&oi=sra, **2012**.

Manuscript received: July 27, 2021
Revised manuscript received: September 20, 2021
Accepted manuscript online: September 21, 2021
Version of record online: ■■, ■■■

FULL PAPERS

Scent glands of male *Heliconius erato* contain up to 350 components with individual compositions. Linear terpenes dominate, either as esters or as long, regular linear isoprene-oligomers. The high chemical diversity is reflected within the many different mimicry patterns of these butterflies. The compounds were identified by a combination of various methods, including synthesis, which allowed to unravel the underlying biosynthetic pathways used to arrive at these compound bouquets.