Comparative metabolomics with Metaboseek reveals functions of a conserved fat metabolism pathway in *C. elegans*

Maximilian J. Helf[1,#], Bennett W. Fox[1,#], Alexander B. Artyukhin[2], Ying K. Zhang[1], Frank C. Schroeder[1,*]

[1]Boyce Thompson Institute and Department of Chemistry and Chemical Biology, Cornell University, Ithaca, New York 14853, United States

[2]Chemistry Department, College of Environmental Science and Forestry, State University of New York, Syracuse, New York 13210, United States

[#]These authors contributed equally to this work

*Correspondence to fs31@cornell.edu

ABSTRACT

Untargeted metabolomics via high-resolution mass spectrometry (HRMS) can reveal more than 100,000 molecular features in a single sample, many of which may represent unidentified metabolites, posing significant challenges to data analysis. We here introduce Metaboseek, an open-source analysis platform designed for untargeted comparative metabolomics and demonstrate its utility by uncovering biosynthetic functions of a conserved fat metabolism pathway, α-oxidation, using *C. elegans* as a model. Metaboseek integrates modules for molecular feature detection, statistics, molecular formula prediction, and fragmentation (MS/MS) analysis, which uncovered more than 200 previously uncharacterized α-oxidation-dependent metabolites in an untargeted comparison of wildtype and α-oxidation-defective *hacl-1* mutants. The identified metabolites support the predicted enzymatic function of HACL-1 and revealed that α-oxidation participates in metabolism of endogenous β-methyl-branched fatty acids and food-derived cyclopropane lipids. Our results showcase compound discovery and feature annotation at scale via untargeted comparative metabolomics applied to a conserved primary metabolic pathway and suggest a model for the metabolism of cyclopropane lipids.

MAIN

Widespread adoption of high-resolution mass spectrometry (HRMS) for untargeted metabolomics has revealed a vast universe of biogenic small molecules, including a large number of compounds whose chemical structures have not been elucidated ("unknowns")[1]. Many of these metabolites may serve important biological functions, as intra- or intercellular signaling molecules, e.g., as hormones, or mediating communication at the inter-organismal level, e.g., in host-microbe interactions or as pheromones[2–5]. Their large-scale identification, quantitation, and elucidation of underlying biosynthetic networks promises to advance mechanistic understanding of phenotypes and complement transcriptomics and proteomics[6–9].

However, the highly irregular and often unpredictable structures of metabolites pose a largely unmet challenge to their systematic chemical and biological annotation. High-performance liquid chromatography (HPLC)-HRMS analysis of a typical metabolome sample of plant or animal origin can reveal more than 100,000 molecular features (defined by a mass-to-charge ratio, *m/z*, and retention time, RT), representing a complex mixture of ions derived from known and unknown metabolites, adducts, naturally occurring isotopes and background[9–13]. Comparative analysis of samples representing different biological conditions or genetic

backgrounds can identify molecular features that are significantly differential between conditions, akin to finding induced or repressed genes in transcriptomics. In the case of untargeted metabolomics, such comparative analyses provide the basis for prioritizing among the many detected unknowns for subsequent structure elucidation[14], which is often time- and resource-intensive.

Processing of metabolomics data involves three major steps – feature detection, comparative statistical analysis, and structural characterization – each of which comes with its own challenges addressed by free, open-source computational tools as summarized in **Supplementary Table S1**. Importantly, comparative metabolomics requires effective, multi-layered interaction with MS and MS fragmentation (MS/MS) data to enable culling and prioritization of differentially regulated molecular features, in particular when the analysis is discovery-oriented and focused on the identification of unknowns. Exploring the metabolomes of *C. elegans* and other model systems, we recognized the need for an open-source analysis platform that can serve as a flexible and customizable hub integrating diverse existing tools.

We here introduce *Metaboseek*, a modular software platform that provides a comprehensive data analysis workflow, from feature detection to compound identification, specifically designed to facilitate untargeted metabolomics. Metaboseek incorporates popular metabolomics tools and makes them available in an intuitive, browser-based graphical user interface (**Figure 1,** also see Metaboseek.com). We then leverage Metaboseek to investigate peroxisomal α-oxidation (pαo) in *C. elegans*, a conserved fatty acid degradation pathway that functions in parallel with β-oxidation and is required for the breakdown of β-branched fatty acids. Whereas β-oxidation takes place in both mitochondria and peroxisomes, α-oxidation occurs exclusively in peroxisomes, which are membrane-bound metabolic compartments that function coordinately with other organelles in lipid and bile acid metabolism[15]. In contrast to the well-studied peroxisomal β-oxidation pathway[16], the pαo pathway in *C. elegans* has not been investigated. Using Metaboseek for untargeted comparative metabolomics of a pαo mutant, we show that disruption of *C. elegans* pαo results in accumulation of several 100 metabolites, most of which have not previously been reported. The identified metabolites support homology-based annotation of the pαo pathway enzyme, HACL-1, and indicate a role for pαo in the processing of bacteria-derived cyclopropane fatty acids.

RESULTS

The Metaboseek workflow. In a typical metabolomics workflow, HPLC-HRMS and MS/MS data acquisition is followed by *Feature Detection* (a “feature” representing a specific *m/z*, RT

78 pair) and *Feature Grouping* (aiming to recognize the same features across different samples, a
79 non-trivial step, especially in the case of closely eluting isomers). Metaboseek integrates the
80 *XCMS* package for feature detection and grouping, generating a feature table containing the
81 data for all identified molecular features in each sample. Alternatively, feature tables can be
82 imported from external tools, such as MZmine[17], MS-DIAL[18,19], or XCMSOnline[20]. Following
83 feature detection and grouping, any subsequent statistical or data-dependent analysis
84 generates new columns in the *Feature Table*, which serves as a customizable information hub
85 that guides data analysis.

86 The *Data Explorer* section is at the heart of Metaboseek, providing the platform to further
87 process, prioritize, sort, and visualize molecular features within the *Feature Table*. An extensive
88 range of filters can be used in any user-defined configuration to enable intuitive prioritization of
89 features, e.g., mean intensity threshold and fold change of a feature between different sample
90 groups, at a specific significance level. Finding molecular features with significantly differential
91 abundances in Metaboseek is assisted by built-in *t*-test and ANOVA, as well as group-wise fold-
92 change metrics. For more advanced statistical analysis, feature tables can be exported directly
93 in a MetaboAnalyst-compatible format[21]. Features of interest can be manually inspected in the
94 interactive user interface, which includes a customizable *Data Viewer* for visual validation of
95 molecular features, modules for *Molecular Formula Prediction* and isotope/adduct assignments,
96 e.g., via CAMERA[22], and removal of background-derived features via *Peak Quality* analysis
97 function, wherein features are scored based on fitting to an idealized peak shape. Following
98 validation, Metaboseek can export annotated feature tables as inclusion lists to facilitate
99 targeted MS/MS data acquisition. MS/MS spectra can be easily compared to each other using
100 the *Keep and Compare* and *Find Patterns* functions (vide infra), or annotated using SIRIUS
101 fragmentation trees and CSI:FingerID database matching[23,24].

102 The *Molecular Networking* tool in Metaboseek uses feature grouping information from
103 *XCMS* to match MS/MS spectra with a corresponding molecular feature. In molecular
104 networking, MS/MS spectra of different features are compared and ranked by similarity of
105 fragmentation patterns[25,26]. Similarity scores are based on peak matching between pairs of
106 averaged spectra, and subsequent calculation of the cosine score between relative peak
107 intensities of the matched spectra, similar to the GNPS feature-based networking workflow[27].
108 The resulting networks, including spectra of individual features, can be easily viewed and
109 evaluated in Metaboseek, which allows users to click nodes and compare MS/MS fragmentation
110 patterns directly. Any information in the feature table, e.g., relative abundances of compounds

111 across sample groups, comments, or statistical analyses, can be mapped onto the network
112 view. Furthermore, the MS/MS networking parameters can be modified using the *Simplify*
113 *Network* function, e.g., by adjusting similarity thresholds (edges), restricting the number of
114 edges per node, or limiting the number of nodes per cluster (**Supplementary Figure 1**).

115 The Metaboseek *Data Explorer* further incorporates an isotope tracing module, *Label*
116 *Finder*, providing an integrated analysis option for stable isotope labeling experiments (**Figure**
117 **1**). Like many popular open-source data analysis tools, Metaboseek is written in R, using the
118 *shiny* R package for interactive data visualization, and thus can be run either on a server or
119 locally on any computer. All data analysis steps are tracked so that settings can be archived and
120 reproduced. Installation files and extensive documentation are available online at
121 Metaboseek.com.

122 **Citronellic acid as a probe for α-oxidation.** A putative *C. elegans* α-oxidation pathway was
123 proposed more than two decades ago, based on sequence similarity to characterized rat and
124 human enzymes (**Figure 2a**)[28]. In humans, the principal function of α-oxidation is presumed to
125 be the enzymatic digestion of β-branched fatty acids (**1**), such as (*R*)-citronellic acid (CA, **6**)
126 (**Figure 2d**), which cannot be processed by β-oxidation[29]. In the first step of the human α-
127 oxidation pathway, fatty acyl-coenzyme A (CoA) derivatives (**2**) are *syn*-hydroxylated at the α-
128 position by phytanoyl-CoA dioxygenase, PHYH, an iron-coordinating enzyme that is 52% and
129 59% identical to the uncharacterized *C. elegans* proteins, ZK550.5 and ZK550.6, respectively
130 (**Figure 2a**)[30]. Genetic mutations in *PHYH* cause Refsum's syndrome, which is characterized by
131 toxic accumulation of branched fatty acids in the blood and nervous system[31]. The next step is
132 catalyzed by 2-hydroxyacyl-CoA lyase, HACL1, which binds thiamine pyrophosphate (TPP) as a
133 cofactor and cleaves a C-C bond in the α-hydroxy, β-methylacyl-CoA (**3**) to produce formyl-CoA
134 and an α-methyl fatty aldehyde (**4**). The *C. elegans* gene *B0334.3*, herein referred to as *hacl-1*,
135 encodes an enzyme 49% identical to human HACL1, including high homology in the TPP-
136 binding domain (**Supplementary Figure 2**). Finally, the aldehyde (**4**) is oxidized in an NAD^+-
137 dependent reaction to the corresponding α-methyl fatty acid (**5**), which now is a suitable
138 substrate for further processing via β-oxidation (**Figure 2a**).

139 To probe p*α*o in *C. elegans*, we compared metabolism of supplemented CA (**6**) in WT
140 animals and *hacl-1(tm6725)* mutants (**Figure 2b**), which harbor a deletion predicted to disrupt
141 the splice acceptor site of the largest exon (**Figure 2c**). *hacl-1(tm6725)* mutants develop
142 normally and exhibit no overt abnormalities under laboratory conditions. If *hacl-1* were required
143 for α-oxidation, supplementation with CA should result in accumulation of an α-hydroxyl CA

144 derivative (**7**), whose MS/MS spectrum should show a characteristic neutral loss of formic acid
145 (**Figure 2d**)[32]. The custom filtering options and extracted ion chromatogram (EIC) display in
146 Metaboseek facilitated screening molecular ions at the expected *m/z* of **7** (*m/z* 185.1183,
147 $C_{10}H_{17}O_3^-$) that produce neutral loss of formic acid during MS/MS and were strongly enriched in
148 *hacl-1* samples (**Figure 2e**). Of six other CA-derived features detected at *m/z* 185.1183, none
149 were enriched in *hacl-1* relative to WT, nor did these features exhibit neutral loss of formic acid
150 in MS/MS, suggesting these features are derived from hydroxylation of CA via other metabolic
151 pathways (**Supplementary Figure 3**).

152 Additional CA-dependent features enriched in *hacl-1* relative to CA-supplemented WT
153 animals were uncovered using a combination of fold change, intensity, and CA-supplementation
154 dependent filters, which, after removal of adducts, revealed 32 CA-derived metabolites
155 (**Supplementary Figure 3**). These differential features included multiply oxygenated CA
156 derivatives, which could be ostensibly derived from ω-oxidation following stalled α-oxidation[33].
157 For example, we detected several features at *m/z* 201.1132 that elute over a wide RT range,
158 representing dihydroxylated CA derivatives, two of which were *hacl-1*-enriched (such as **8**, see
159 also **Supplementary Figure 3**). Additional *hacl-1*-dependent CA-derivatives included putative
160 ethanolamides (**9**), glycosides (**10**), and an *N*-acyl glycerophosphoethanolamide conjugate
161 (**Supplementary Figure 3**). For most of these compounds, MS/MS fragmentation between the
162 α- and carbonyl carbons suggested α-hydroxylation of the citronellyl moiety (**Figure 2d**). Similar
163 results were obtained when *hacl-1* worms were supplemented with geranic or phytanic acid
164 (data not shown). Taken together, this supplementation study revealed a set of *hacl-1*-
165 dependent shunt metabolites of CA, consistent with the proposed function of HACL-1 as a 2-
166 hydroxyacyl-CoA lyase.

167 **Endogenous C_{11} fatty acids enriched in *hacl-1* larvae.** We next investigated the impact of
168 *hacl-1* inactivation on endogenous metabolites, using molecular networking of MS/MS spectra
169 acquired as part of in-depth untargeted HPLC-HRMS analysis of *hacl-1* mutants and WT *C.*
170 *elegans*. To assess the role of *hacl-1* in *C. elegans* metabolism, we initially focused on starved
171 animals at the first larval stage (L1), a condition that allows the study of *C. elegans* metabolism
172 in the absence of bacterial food[34]. Comparative analysis of negative ion MS data for L1 larvae
173 revealed a small set of features strongly enriched in *hacl-1* relative to WT, several of which
174 clustered together in the MS/MS network (**Figure 3a**). Stringent fold change (10-fold), intensity
175 (top 1.25% of detected features), and unadjusted significance ($p < 0.05$) thresholds yielded 57
176 molecular features that were highly enriched in *hacl-1* larvae. Following *CAMERA*

177 isotope/adduct assignment and manual curation, we detected 14 *hacl-1*-dependent metabolites
178 (**Figure 3b**). Intriguingly, the majority of these compounds appeared to represent C_{11} fatty acids,
179 based on their ionization properties and MS/MS spectra. Furthermore, MS/MS spectra of the
180 most abundant *hacl-1*-enriched metabolites featured a product ion with *m/z* 72.993
181 corresponding to glyoxylate ($C_2HO_3^-$), which was not observed in any CA-derived metabolites
182 (**Figure 3c**).

183 We selected one highly abundant *hacl-1*-enriched metabolite ($C_{11}H_{19}O_4^-$) for isolation by
184 preparative HPLC followed by structure elucidation via 2D NMR spectroscopy (**Supplementary**
185 **Figure 4** and **Supplementary Table S2**), which revealed an unusual β-methyl substituted,
186 eleven-carbon fatty acid, named bemeth#3 (**11**, **Figure 3c**). α-Hydroxylation and the position of
187 the double bond in bemeth#3 are consistent with the strong glyoxylate product ion in its MS/MS
188 spectrum. Moreover, the structure of bemeth#3 suggested that other glyoxylate ion-producing
189 metabolites that accumulate in *hacl-1* mutants also represent derivatives of α-hydroxylated β-
190 methyl-4-decenoic acid (e.g., **12**, **Figure 3c**). These assignments were further supported via
191 synthesis of an authentic sample of the two diastereomers of **12** (**Figure 3d** and
192 **Supplementary Table S3**), whose MS/MS spectra and retention times were identical to those
193 of the corresponding *hacl-1*-enriched metabolites (**Figure 3e**). Other metabolites enriched in
194 *hacl-1* include less abundant isomers of **11** and **12** with identical MS/MS fragmentation, as well
195 as derivatives that appear to have undergone additional oxidation, including putative
196 dicarboxylic acids, such as $C_{11}H_{17}O_5^-$ (**Supplementary Figure 5**). Taken together, analysis of
197 *hacl-1* larvae revealed an unusual family of C_{11} fatty acids based on the β-methyl-decenoic acid
198 scaffold, which has not been previously reported from animals. The presence of an α-hydroxyl
199 group in the identified C_{11} acids **11** and **12** suggests that they represent plausible substrates of
200 HACL-1 and therefore accumulate in *hacl-1* mutants. The more highly oxygenated derivatives,
201 such as $C_{11}H_{17}O_5^-$, could result from ω-oxidation of **12** as part of a shunt pathway, similar to the
202 role of ω-oxidation in human fatty acid metabolism[33].

203 **Comparative metabolomics of *hacl-1* adults.** Next, we employed Metaboseek for
204 comparative metabolomics of adult-stage *hacl-1* mutant and WT animals. Conditioned culture
205 medium (*exo*-metabolome) and worm bodies (*endo*-metabolome) were harvested separately,
206 extracted, and analyzed by HPLC-HRMS/MS in positive and negative ionization modes, yielding
207 more than 100,000 features combined following blank subtraction and *Peak Quality* thresholding
208 (see Methods). Like *hacl-1* mutant larvae, *hacl-1* adults accumulate the β-branched C_{11} acid **11**
209 and related metabolites (**Supplementary Figure 6**). However, in contrast to L1 larvae, these

210 C_{11} acid derivatives were not the most differential metabolites in *hacl-1* adults (**Supplementary**
211 **Figure 6**). Untargeted analysis using intensity (top 6.5% of detected features), unadjusted
212 significance ($p < 0.05$), and fold-change (5-fold) filters uncovered >1,000 features that were
213 enriched in the *hacl-1 exo*-metabolome, which we explored by MS/MS networking.

214 The majority of differential features clustered in five major subnetworks (SN1-SN5,
215 **Figure 4**). Inspection of SN1 revealed several homologous series of features, related by the
216 mass difference of a methylene (Δ*m/z* 14.0156). Use of the "*Keep and Compare"* functionality in
217 the *MS/MS Browser* facilitated simultaneous display of multiple MS/MS spectra and
218 automatically highlights fragments shared between spectra, e.g., conserved product ions that
219 correspond to a phosphoethanolamine moiety, a phosphorylated hexose, and ions indicating a
220 phosphate group (**Supplementary Figure 7**). Complementary analysis of MS/MS fragmentation
221 patterns in positive ionization mode further supported that SN1 represents *N*-acyl
222 glycophosphoethanolamides ("*N*-acyl GPEs"), including saturated (N:0), singly unsaturated
223 (N:1), polyunsaturated (N:n), and mono-oxygenated (mN:n) acyl moieties ranging from C_8-C_{20}.
224 In many cases, several isobaric features were detected, e.g., four distinct isomers of *N*-acyl
225 GPE-14:2 (**Supplementary Figure 8**). Enrichment trends were similar across the *endo*- and the
226 *exo*-metabolomes, but *N*-acyl GPE were 10-100-fold more abundant in the *exo*-metabolome
227 (**Supplementary Figure 8**).

228 **Finding MS/MS patterns with Metaboseek.** Analysis of SN2 indicated that it represents *N*-acyl
229 glycoglycerophosphoethanolamides ("GLEA", **Figure 4**). This compound family had been
230 previously described in the context of ethanol-dependent *de novo* fatty acid biosynthesis in
231 starved *C. elegans* larvae[35]. MS/MS spectra of GLEA exhibit a characteristic product ion at *m/z*
232 333.0592, in addition to phosphoglycerol and several glucose-derived fragments, which were
233 used to define a "*Pattern*" in Metaboseek. The *Find Patterns* function in Metaboseek profiles all
234 MS/MS spectra for user-defined fragmentation patterns, including neutral losses; hits are
235 matched and recorded in new interactive columns in the feature table, which enabled rapid
236 identification of more than 100 GLEA-like molecular features distributed across SN2 and two
237 additional subnetworks, SN3 and SN4 (**Figure 4**).

238 Each of the three GLEA clusters revealed slightly different MS/MS fragmentation,
239 providing important structural clues. GLEA in SN3 produced additional product ions with *m/z*
240 376.1016 and 418.1123, suggesting that these metabolites are β-hydroxylated, which then
241 results in fragmentation between the α and β carbons (**Figure 4**). Larger cultures of *hacl-1* were
242 grown and extracted to isolate the most abundant compound from SN3, GLEA-m16:1, one of

243 the most intense and differential features in the entire *exo*-metabolome. 2D NMR spectroscopic
244 analysis established this compound as a 2-*O*-(β-glucosyl)-glycero-1-phosphoethanolamide of β-
245 hydroxylated hexadecenoic acid (**16**, **Supplementary Table S4**). β-hydroxylation of the fatty
246 acyl moiety in this compound is consistent with observed fragmentation between the α and β
247 carbons of the fatty acyl group and suggests that other metabolites in SN3 also represent GLEA
248 of β-hydroxylated fatty acids (**Figure 4**). Upregulation of β-hydroxylated lipid derivatives
249 suggested that mitochondrial or peroxisomal β-oxidation may be perturbed in *hacl-1* mutants.
250 However, production of ascaroside pheromones, which relies on peroxisomal β-oxidation, was
251 largely unchanged in *hacl-1* mutants compared to WT (**Supplementary Figure 9**), suggesting
252 that *hacl-1* inactivation may interact with mitochondrial β-oxidation.

253 GLEA in SN4 did not undergo fragmentation across the α and β carbons, but instead
254 produced an intense product ion with *m/z* 376.1016, corresponding to fragmentation across the
255 amide bond. GLEA in SN4 were much less abundant and eluted later than isobaric metabolites
256 in SN3. We hypothesize that SN4 represents GLEA bearing α-hydroxy acyl substituents;
257 however, their low abundance precluded NMR spectroscopic characterization (**Supplementary**
258 **Figure 10**). Lastly, analysis of SN5 revealed a large family of *hacl-1*-enriched *N*-acyl
259 ethanolamides (NAEs) (**Figure 4**). All NAE in SN5 produced the product ion with *m/z* 102.056,
260 corresponding to cleavage between the α- and β-carbon of the acyl group, suggesting β-
261 hydroxylation in analogy to SN3.

262 **Stable isotope tracing with Metaboseek.** Comparing the series of *N*-acyl-GPEs, GLEA, and
263 NAEs enriched in *hacl-1* worms, we noted that derivatives of mono-unsaturated C_{13}- and C_{15}-
264 fatty acids were among the most abundant *hacl-1*-enriched compounds, even though the
265 corresponding free fatty acids are not particularly abundant in *C. elegans*[36]. Generally, odd-
266 chain fatty acids in *C. elegans* are derived primarily either from iso-branched chain fatty acid
267 (BCFA) biosynthesis, which employs leucine-derived isovaleryl-CoA as a starter unit[37], or,
268 alternatively, from metabolism of diet-derived cyclopropane fatty acids, which are abundantly
269 produced by the bacterial diet, *E. coli* OP50[38].

270 We first tested whether the major *hacl-1*-enriched monounsaturated C_{13} and C_{15} lipids
271 are derived from BCFA metabolism. For this purpose, we grew worms supplemented with $^{13}C_6$-
272 labeled leucine, which we reasoned should result in $^{13}C_5$-enrichment of BCFAs and any derived
273 *N*-acyl GPE, GLEA, and NAE (**Figure 5a**). The *Label Finder* tool in Metaboseek facilitated
274 profiling $^{13}C_5$- and $^{13}C_6$-enrichment for discovery of BCFA- and Leu-derived metabolites. This
275 analysis revealed several hundred isotope-enriched features, including iso-branched fatty acids

276 and derivatives thereof, which were visually validated using the *Mass Shifts* feature in
277 Metaboseek to display EICs corresponding to incorporation of $^{13}C_5$ (Δ *m/z*, 5.0167, **Figure 5b**).
278 However, the most abundant *hacl-1*-enriched compounds harboring 13:1 and 15:1 acyl groups
279 showed no evidence for label incorporation, indicating that these unsaturated odd chain lipids
280 do not originate from BCFA metabolism (**Supplementary Figure 11**).

281 **Cyclopropane fatty acids accumulate in *hacl-1.*** We then asked whether the
282 monounsaturated C_{13} and C_{15} fatty acyl derivatives accumulating in *hacl-1* mutants are derived
283 from bacterial cyclopropane fatty acids. In the case of *E. coli* OP50, C_{17} and C_{19} cyclopropane
284 lipids can account for nearly half of all lipid species and thus comprise a substantial portion of *C.*
285 *elegans* lipid intake[39]. To test whether the monounsaturated C_{13} and C_{15} lipids enriched in *hacl-1*
286 are derived from cyclopropane fatty acids, we compared the metabolomes of worms fed either
287 OP50 or JW1653-1 bacteria, a cyclopropane-deficient *E. coli* strain[40] (**Figure 5c**). First, we
288 confirmed via 2D NMR spectroscopy that JW1653-1 does not produce cyclopropane lipids and
289 that worms fed JW1653-1 bacteria do not produce cyclopropane lipids (**Supplementary Figure**
290 **12**). Next, we compared the metabolomes of animals grown on OP50 or JW1653-1 via HPLC-
291 HRMS, which revealed that production of the most abundant *hacl-1*-enriched *N*-acyl GPEs was
292 abolished in JW1653-1-fed worms (**Figure 5d**). Additional *N*-acyl GPEs enriched in *hacl-1*
293 mutant were also found to be dependent on bacterial cyclopropane fatty acid biosynthesis,
294 including multiple hydroxylated *N*-acyl GPE species (**Figure 5e**). Untargeted comparative
295 analysis of OP50- and JW1653-1-fed worms revealed a large number of other cyclopropane-
296 containing metabolites, including GLEA, as well as putative oxidized fatty acids and fatty acyl
297 glycosides, many of which also accumulate in *hacl-1* worms (**Figure 5f** and **Supplementary**
298 **Figure 13**). Taken together, our results suggest that diet-derived C_{17} or C_{19} cyclopropane fatty
299 acids are initially chain shortened via β-oxidation to yield shorter chained derivatives that
300 become substrates for pαo. If pαo is blocked, as in the case of *hacl-1* inactivation, β-oxidation
301 intermediates are shunted towards production of, e.g., *N*-acyl GPE, GLEA, and other lipids
302 (**Figure 5g**).

303

304 **DISCUSSION**

305 We here demonstrated the use of Metaboseek for a multi-layered comparative metabolomics
306 study of a conserved fatty acid metabolism pathway, pαo, in *C. elegans*. By probing metabolism
307 of WT and *hacl-1* mutants with a pαo test substrate, CA, we confirmed the predicted enzymatic
308 function of HACL-1 as a 2-hydroxyacyl-CoA lyase. Subsequent untargeted comparison revealed

309 pervasive changes in lipid metabolism in *hacl-1* mutants, including accumulation of an unusual
310 family of α-hydroxylated β-branched C_{11} acids. Their abundant production and life stage-specific
311 regulation suggests that β-branched C_{11} acids – perhaps a precursor or downstream metabolite
312 of **11** and **12** – may serve specific functions in *C. elegans*. In addition to the C_{11} acids, *hacl-1*
313 mutants accumulate an unexpected diversity of modular lipids derived from the intersection of
314 multiple branches of fatty acid metabolism with NAE biosynthesis. Particularly abundant among
315 lipids accumulating in *hacl-1* are derivatives of cyclopropyl fatty acids, suggesting that pαo
316 participates in cyclopropyl metabolism. Mechanisms for the breakdown of cyclopropyl fatty acids
317 have remained largely unknown, though cyclopropyl lipids have previously been shown to affect
318 recovery from larval diapause (dauer)[41]. Cyclopropyl lipids are also present in the human diet,
319 most prominently in cheese and dairy originating from animals fed fermented grains; however,
320 whether α-oxidation plays a role in the mammalian metabolism of cyclopropyl lipids remains to
321 be determined[42,43].

322 Comparative analysis with Metaboseek revealed a large number of additional
323 differences between the metabolomes of WT and *hacl-1* animals, of which many represent
324 previously undescribed metabolites (**Supplementary Table S5**). For all newly annotated
325 metabolites, this table includes retention time, *m/z*, putative molecular formulae and compound
326 class assignments, fold-change, as well as isotopic enrichment data, which will facilitate follow-
327 up studies in conjunction with the deposited MS raw data[44–46]. As we here demonstrated, in-
328 depth evaluation of MS and MS/MS raw data is key to discovery-oriented workflows. Online
329 resources such as GNPS and MassBank provide access to vast amounts of MS data,
330 highlighting the need for versatile tools that facilitate raw data analysis for metabolomics[47–49].
331 For this purpose, Metaboseek combines more than 60 different modules which tool developers
332 can use as building blocks for specialized data analysis apps with minimal effort. The seamless
333 integration of intuitive data filters and a range of analysis tools facilitates metabolite annotation
334 up to confidence Level 3 for many detected features, enabling tentative structure or compound
335 class assignments[50–52]. In parallel, facile statistical analysis of metabolite variation across
336 multiple genotypes and/or environmental conditions enables developing functional and
337 biosynthetic hypotheses. Further, the output from Metaboseek facilitates intersecting
338 metabolomics with transcriptomics, proteomics, or genomics, toward a systems-level
339 understanding of biosynthetic networks and metabolite functions. Metaboseek thus provides a
340 flexible and expandable open-source platform to accelerate chemical and biological annotation
341 of metabolites, including the large space of yet unidentified biogenic small molecules.

342 Our comparative metabolomics analysis of α-oxidation shows that even a primary
343 metabolic pathway in an otherwise well-studied model system can reveal a large number of
344 previously uncharacterized compounds, as well as unexpected connections to other pathways,
345 e.g., β-oxidation or cyclopropane fatty acid metabolism. Like much of conserved primary
346 metabolism, α-oxidation was initially characterized more than 50 years ago[53,54]. It seems likely
347 that re-analysis of primary metabolic pathways using state-of-the-art HRMS and data analysis
348 tools will synergize with transcriptomic and proteomic studies to harness the potential of
349 metabolomics as the 'omics discipline that most closely reflects phenotype[7–9].

350

351

352 **METHODS**

353 ***C. elegans* strains.** Unless otherwise indicated, worms were maintained on Nematode Growth
354 Medium (NGM) 6 cm diameter petri dish plates seeded with *E. coli* OP50 obtained from the
355 *Caenorhabditis* Genetics Center (CGC). For experiments with cyclopropane deficient bacteria,
356 worms were grown on NGM 6cm plates seeded with *E. coli* JW1653-1, a kind gift from the
357 Walhout Lab (University of Massachusetts Medical School, Worcester, MA). The following *C.*
358 *elegans* strains were used for comparative metabolomics: Bristol N2 ("wildtype") obtained from
359 the CGC, and *B0334.3(tm6725)* obtained from the National Bioresource Project, Tokyo,
360 Japan[55], referred to as *hacl-1,* strain designation FCS7. The FCS7 strain was the result of
361 backcrossing *tm6725* with *sqt-1 (*FCS6) for seven generations. FCS6 was iteratively
362 backcrossed with Bristol N2 for a total of six generations. After the final backcross, FCS7
363 hermaphrodites were singled and allowed to self, non-rollers were picked, and the genotype
364 was confirmed by PCR and Sanger sequencing.

365 ***C. elegans* liquid cultures.** Analysis of starved L1 larvae followed a previously described
366 procedure[56]. For the analysis of staged gravid adults, approximately 75,000 synchronized L1
367 larvae were added to 125 mL Erlenmeyer flasks containing 25 mL S-complete medium and
368 kanamycin at 35 µg/mL to prevent contamination. Worms were fed with 50x concentrated *E. coli*
369 OP50 or *E. coli* JW1653-1 and incubated at 20 °C with shaking at 180 RPM for 66–70 h, at
370 which time the population was predominantly gravid adults, determined by microscopic
371 inspection. Control samples to account for bacterial matrix were prepared with the same amount
372 of *E. coli* OP50 or JW1653-1 under identical conditions. Liquid cultures were transferred to 50
373 mL conical tubes and centrifuged (500 x g, 22 °C, 1 min), and the top 20 mL of the resulting
374 supernatant (*exo*-metabolome) was transferred to a fresh conical tube and snap frozen.
375 Remaining worm pellet was transferred to a 15 mL conical tube, centrifuged (500 x g, 22 °C, 1
376 minute), and washed three times with M9 before snap freezing in liquid nitrogen.

377 **Test substrate feeding experiments.** Approximately 100,000 synchronized L1 larvae were
378 added to 125 mL Erlenmeyer flasks containing 10 mL M9 media and 300 µM citronellic acid
379 (Sigma-Aldrich 303429), phytanic acid (Sigma-Aldrich P4060), retinoic acid (Sigma-Aldrich
380 R2625), geranic acid (Aldrich 427764), or an equivalent volume of methanol only (vehicle
381 control) and were incubated at 20 °C with shaking at 180 RPM for 24 hrs. Cultures were
382 transferred to 15 mL conical tubes and centrifuged (500 x g, 22 °C, 1 min), and the resulting
383 supernatant (*exo*-metabolome) was transferred to a fresh conical tube and snap frozen.
384 Remaining L1 pellet was washed three times with M9 before snap freezing in liquid nitrogen.

385 **$^{13}C_6$-Leu isotope tracing experiment.** Approximately 60,000 synchronized N2 (WT) L1 larvae
386 were added to 125 mL Erlenmeyer flasks containing 20 mL S-Complete medium. Worms were
387 fed with 60 mg freeze-dried OP50 powder (InVivoBiosystems, formerly NemaMetrix Inc., OP-50-
388 31772) and supplemented with leucine (Sigma Aldrich L8000) or $^{13}C_6$-leucine (Cambridge
389 Isotope Laboratories CLM-2262-H-PK) at a final concentration of 2 mM. Worms were incubated
390 at 20 °C with shaking at 180 RPM for 66–70 hrs, at which time the population was a mixture of
391 young and gravid adults, determined by microscopic inspection. Liquid cultures were centrifuged
392 (500 x g, 22 °C, 1 min), and the resulting supernatant was snap frozen. Worm pellet was
393 washed three times with M9 before snap freezing in liquid nitrogen.

394 **Sample preparation for HPLC-MS.** *Exo*-metabolome (conditioned media) samples were
395 lyophilized ~48 hrs using a VirTis BenchTop 4K Freeze Dryer. Dried material was directly
396 extracted in 10 mL methanol in 20 mL glass vials stirred overnight. Vials were centrifuged at
397 2750 RCF for five minutes in an Eppendorf 5702 Centrifuge using rotor F-35-30-17. The
398 resulting supernatant was transferred to a clean 20 mL glass vial and concentrated to dryness in
399 an SC250EXP Speedvac Concentrator coupled to an RVT5105 Refrigerated Vapor Trap
400 (Thermo Scientific). The resulting powder was suspended in methanol and analyzed directly by
401 HPLC-MS, as described below. *Endo*-metabolome (nematode bodies) were lyophilized for 18-
402 24 hrs using a VirTis BenchTop 4K Freeze Dryer. Dried pellets were transferred to 1.5 mL
403 microfuge tubes and disrupted in a Spex 1600 MiniG tissue grinder after the addition of two
404 stainless steel grinding balls to each sample. Microfuge tubes were placed in a Cryoblock
405 (Model 1660) cooled in liquid nitrogen, and samples were disrupted at 1100 RPM for 60 s. This
406 process was repeated two additional rounds for a total of three disruptions. Pellets were
407 transferred to 8 mL glass vials in 5 mL methanol and stirred overnight. Subsequent steps for
408 concentration and resuspension were followed as described for the *exo*-metablome.

409 **Mass spectrometry.** Liquid chromatography was performed on a Vanquish HPLC system
410 controlled by Chromeleon Software (ThermoFisher Scientific) and coupled to an Orbitrap Q-
411 Exactive High Field mass spectrometer controlled by Xcalibur software (ThermoFisher
412 Scientific). Methanolic extracts prepared as described above were separated on a Thermo
413 Hypersil Gold C18 column (150 mm x 2.1 mm, particle size 1.9 µM; 25002-152130) maintained
414 at 40 °C with a flow rate of 0.5 mL/min. Solvent A: 0.1% formic acid (Fisher Chemical Optima
415 LC/MS grade; A11750) in water (Fisher Chemical Optima LC/MS grade; W6-4); solvent B: 0.1%
416 formic acid in acetonitrile (Fisher Chemical Optima LC/MS grade; A955-4). A/B gradient started
417 at 1% B for 3 min after injection and increased linearly to 98% B at 20 min, followed by 5 min at

418 98% B, then back to 1% B over .1 min and finally held at 1% B for the remaining 2.9 min to re-
419 equilibrate the column (28 min total method time). Mass spectrometer parameters: spray
420 voltage, -3.0 kV / +3.5 kV; capillary temperature 380 °C; probe heater temperature 400 °C;
421 sheath, auxiliary, and sweep gas, 60, 20, and 2 AU, respectively; S-Lens RF level, 50;
422 resolution, 120,000 at *m/z* 200; AGC target, 3E6. Each sample was analyzed in negative (ESI-)
423 and positive (ESI+) electrospray ionization modes with *m/z* range 100-1000. Parameters for
424 MS/MS (dd-MS2): MS1 resolution, 60,000; AGC Target, 1E6. MS2 resolution, 30,000; AGC
425 Target, 2E5. Maximum injection time, 60 msec; Isolation window, 1.0 *m/z*; stepped normalized
426 collision energy (NCE) 10, 30; dynamic exclusion, 5 sec; top 8 masses selected for MS/MS per
427 scan. Inclusion lists with 20 sec windows were generated in Metaboseek for targeted MS/MS.

428 **Metaboseek analysis.** HPLC-MS data were analyzed using Metaboseek software
429 (documentation available at Metaboseek.com) after conversion to mzXML file format using
430 MSConvert (version 3.0, ProteoWizard[57]); for a full list of supported file types, see section 3.3.4
431 *Supported File Types*. A subset of the mzXML files used in this study are provided as an
432 example data set included with the Metaboseek download at Metaboseek.com. The authors
433 recommend installing and running the software locally. For large datasets, 32 GB memory and
434 modern processor (Intel core i7 7700 / Ryzen 7 1700 or better) is recommended. For analysis
435 with up to 50 files (28-minute method), 16 GB of memory are usually enough. Following
436 conversion to mzXML, data were analyzed using the XCMS-module within Metaboseek with
437 default settings, as described in section 3.5 *XCMS Analysis*. Peak detection was carried out
438 with the *centWave* algorithm using the "Metaboseek_default" settings: 4 ppm, 3_20 peakwidth,
439 3 snthresh, 3_100 prefilter, FALSE fitgauss, 1 integrate, TRUE firstBaselineCheck, 0 noise,
440 wMean mzCenterFun, −0.005 mzdiff. Default settings for XCMS feature grouping: 0.2 minfrac, 2
441 bw, 0.002 mzwid, 500 max, 1 minsamp, FALSE usegroups. Metaboseek peak filling used the
442 following settings: 3 ppm_m, 3 rtw, TRUE rtrange, FALSE areaMode. The XCMS-generated
443 feature table was loaded into the Metaboseek *Data Explorer* along with relevant MS files by
444 designating a project folder, see section 3.3.2 *Load a Metaboseek Project Folder*. MS data
445 display and MS data table were grouped according to genotype and experimental condition via
446 *Regroup MS data* (section 3.4.2) and *Regroup Table* (section 3.4.4.4), respectively. After
447 defining groups as i) *C. elegans* samples, ii) bacterial matrix samples, or iii) blanks, blank
448 subtraction was performed such that any feature less than ten-fold more abundant in *C. elegans*
449 samples than in blanks was removed. The resulting feature list was further culled using the *Fast*
450 *Peak Shapes* (Peak Quality) analysis (**Supplementary Table S6**). These settings were selected

451 for discovery-oriented comparative metabolomics, aiming to retain as many likely “real” features
452 as possible. The resulting table was then regrouped according to *C. elegans* genotype (see
453 *Regroup Table*, section 3.4.4.4) and *Basic Analysis* was performed with WT as the “control”
454 group; see section 3.4.4.3 *Analyze Table* for a complete list of analysis and normalization
455 options. Following analysis, relevant MS/MS data were loaded as described in section 3.3.1
456 *Load MS Data Files Directly*, followed by matching of MS/MS scans to the MS1 files as
457 described in section 3.4.4.3.3 *Advanced Analysis* under the subheading *Find MS2 scans*. At this
458 point, the table was culled to include only MS/MS-matched features, then culled further using a
459 retention time filter, as described in section 3.4.4 *Feature Table Actions*. MS/MS networking was
460 performed as described in section 3.4.2.2 *Compare MS2,* see also **Supplementary Figure 1** for
461 an example of the *Simplify Network* modifications and display options. All plots and reports,
462 including SIRIUS fragmentation trees, can be exported as vector graphics for rapid and efficient
463 sharing, outlined in section 3.4.2.2 *Feature Report*. HRMS data were analyzed using
464 "Metaboseek_default” settings and normalized to the abundance of ascr#3. The identities of all
465 quantified metabolites were verified by analysis of MS/MS spectra and/or synthetic standards.

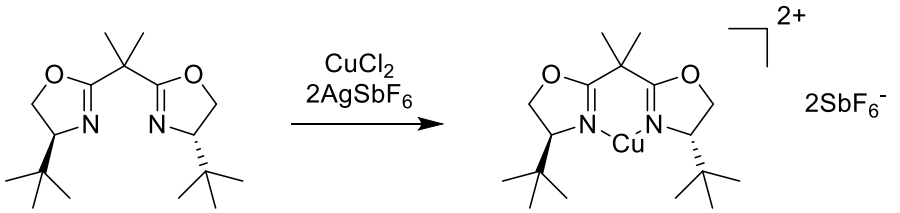
466 **Isolation and NMR spectroscopy of bemeth#3 (11) and GLEA-m16:1 (16).** For isolation of
467 GLEA-m16:1 (**16**) from adult *exo*-metabolome, conditioned media from several medium scale *C.*
468 *elegans hacl-1* cultures was lyophilized and extracted with methanol (as described). Dried
469 methanol extract was loaded onto Celite and fractionated using medium pressure reverse phase
470 chromatography (15 g C18 Combiflash RediSep column, Teledyne Isco 69-2203-334). Solvent
471 A: 0.1% acetic acid in water; solvent B: acetonitrile. Column was primed with 1% B; separation
472 was achieved by 5% B for 2 column volumes (CV), which was increased linearly to 50% B over
473 15 CV, then to 100% B over 3 CV and held at 100% B for 5 CV, before returning to 80% B for 3
474 CV. Fractions were assayed for compounds of interest by HPLC-MS, the relevant fractions were
475 combined and dried *in vacuo*. Following suspension in water: methanol (1:2), the pooled
476 fractions were further separated by semi-preparative HPLC on a Thermo Hypersil Gold C18
477 column (250 mm x 10 mm, particle size 5 µM; 25005-259070) using a Vanquish UPLC system
478 controlled by Chromeleon Software (ThermoFisher Scientific) and coupled to a Dionex UltiMate
479 3000 Automated fraction collector and to an Orbitrap Q-Exactive High Field mass spectrometer
480 using a 9:1 split. Fractions containing GLEA-m16:1 were combined and analyzed by 2D NMR
481 spectroscopy (CD_3OD, Bruker AVANCE III HD, 800 MHz). For spectroscopic data, see
482 **Supplementary Table S4**. For isolation of bemeth#3, conditioned media from several starved
483 L1 cultures was extracted and fractionated analogously. For spectroscopic data, see

484 **Supplementary Table S2**, and relevant section of the dqfCOSY spectrum are shown in
485 **Supplementary Figure 4**.

486 **Chemical syntheses**

487 **General synthetic procedures.** All oxygen and moisture-sensitive reactions were carried out
488 under argon atmosphere in flame-dried glassware. Solutions and solvents sensitive to moisture
489 and oxygen were transferred via standard syringe and cannula techniques. All commercial
490 reagents were purchased as reagent grade and, unless otherwise stated, were purchased from
491 Sigma-Aldrich and used without any further purification. Acetic acid (AcOH), acetonitrile (ACN),
492 dichloromethane (DCM), ethyl acetate (EtOAc), formic acid, hexanes and methanol (MeOH)
493 used for chromatography and as a reagent or solvent were purchased from Fisher Scientific.
494 Thin-layer chromatography (TLC) was performed using J. T. Baker Silica Gel IB2F plates. Flash
495 chromatography was performed using Teledyne Isco CombiFlash systems and Teledyne Isco
496 RediSep Rf silica and C18 columns. All deuterated solvents were purchased from Cambridge
497 Isotopes. Nuclear Magnetic Resonance (NMR) spectra were recorded on Bruker INOVA 500
498 (500 MHz) and Varian INOVA 600 (600 MHz) spectrometers at Cornell University's NMR facility
499 and Bruker AVANCE III HD 800 MHz (800 MHz) or Bruker AVANCE III HD 600 MHz (600 MHz)
500 at SUNY ESF's NMR facility. ^{1}H NMR chemical shifts are reported in ppm (δ) relative to residual
501 solvent peaks (7.26 ppm for chloroform-d, 3.31 ppm for methanol-d_4). NMR-spectroscopic data
502 are reported as follows: chemical shift, multiplicity (s = singlet, d = doublet, t = triplet, q =
503 quartet, m = multiplet, br = broad), coupling constants (Hz). ^{13}C NMR chemical shifts are
504 reported in ppm (δ) relative to residual solvent peaks (77.16 ppm for chloroform-d, 49.00 ppm
505 for methanol-d_4). All NMR data processing was done using MNOVA 14.2.1
506 (https://mestrelab.com/).

507 **Preparation of [Cu(*S,S*)-*tert*-butylbis(oxazolinyl)](SbF$_6$)$_2$**



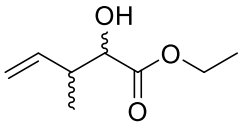
[Cu(S,S)-tert-butylbis(oxazolinyl)](SbF$_6$)$_2$

508

509 To a solution of (*S,S*)-2,2′-isopropylidene-bis(4-tert-butyl-2-oxazoline) (200 mg, 0.68 mmol) in
510 dry dichloromethane (8 mL), $CuCl_2 \cdot H_2O$ (139 mg, 0.82 mmol) and $AgSbF_6$ (583 mg, 0.7 mmol)

511 were added, and the resulting mixture was stirred at room temperature for 14 hours. The green
512 solution was filtered through cotton for later use.

513 **Ethyl 2-hydroxy-3-methylpent-4-enoate (14)**



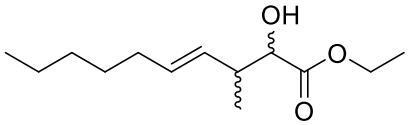
514

515 *cis*-butene was bubbled through a solution of ethyl glyoxylate (2 mL, 9.82 mmol) and [Cu(*S,S*)-
516 *tert*-butylbis(oxazolinyl)]$(SbF_6)_2$ at -78 °C in dry dichloromethane (2 mL), and the reaction was
517 sealed to react at 40 °C. After 2 days, the reaction was concentrated *in vacuo*. Flash column
518 chromatography on silica using a gradient of 0-30% ethyl acetate in hexane afforded **14** as a
519 clear oil (977 mg, 63%, mixture of diastereomers).

520 Major diastereomer, ^{1}H NMR (600 MHz, chloroform-*d*): δ (ppm) 5.74 (ddd, *J* = 17.8, 9.8, 8.1 Hz,
521 1H), 5.07 (m, 1H), 5.05 (m, 1H), 4.24 (m, 2H), 4.10 (dd, *J* = 6.2, 3.4 Hz, 1H), 2.72 (d, *J* = 6.2 Hz,
522 1H), 2.69-2.61 (m, 1H), 1.29 (t, *J* = 7.2 Hz, 3H), 1.16 (d, *J* = 7.0 Hz, 3H).

523 Minor diastereomer, ^{1}H NMR (600 MHz, chloroform-*d*): δ (ppm) 5.85 (ddd, *J* = 17.4, 10.3, 7.4
524 Hz, 1H), 5.12 (m, 1H), 5.09 (m, 1H), 4.26 (m, 2H), 4.16 (dd, *J* = 6.3, 3.7 Hz, 1H), 2.76 (d, *J* = 6.3
525 Hz, 1H), 2.69-2.61 (m, 1H), 1.29 (t, *J* = 7.2 Hz, 3H), 1.01 (d, *J* = 7.0 Hz, 3H).

526 **Ethyl (*E*)-2-hydroxy-3-methyldec-4-enoate (15)**



527

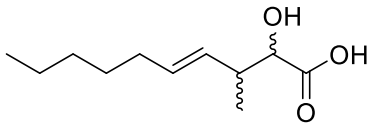
528 To a solution of **14** (950 mg, 6.00 mmol) dissolved in dry dichloromethane (20 mL), Grubbs
529 catalyst 2nd generation (94 mg, 0.15 mmol) and 1-heptene (2.52 mL, 18.00 mmol) were added
530 at room temperature. The reaction mixture was allowed to stir 20 h and was concentrated *in*
531 *vacuo*. Flash column chromatography on silica using a gradient of 0-30 % ethyl acetate in
532 hexanes afforded **15** as a clear oil (785 mg, 76% BRSM, mixture of diastereomers).

533 Major diastereomer, ^{1}H NMR (600 MHz, chloroform-*d*): δ (ppm) 5.46 (dt, *J* = 15.4, 6.7 Hz, 1H),
534 5.31 (dd, *J* = 15.4, 8.2 Hz, 1H), 4.27-4.18 (m, 2H), 4.06 (m, 1H), 2.67 (d, *J* = 5.2 Hz, 1H), 2.60

535 (m, 1H), 1.96 (q, *J* = 7.3 Hz, 2H), 1.36-1.20 (m, 6H), 1.29 (t, *J* = 7.3 Hz, 3H), 1.13 (d, *J* = 7.1 Hz,
536 3H), 0.87 (t, *J* = 7.2 Hz, 3H).

537 Minor diastereomer, ^{1}H NMR (600 MHz, chloroform-*d*): δ (ppm) 5.52 (dt, *J* = 15.4, 6.5 Hz, 1H),
538 5.40 (dd, *J* = 15.4, 7.8 Hz, 1H), 4.27-4.18 (m, 2H), 4.09 (m, 1H), 2.70 (d, *J* = 6.8 Hz, 1H), 2.56
539 (m, 1H), 2.00 (q, *J* = 7.2 Hz, 2H), 1.36-1.20 (m, 6H), 1.28 (t, *J* = 7.2 Hz, 3H), 1.12 (d, *J* = 7.0 Hz,
540 3H), 0.87 (t, *J* = 7.1 Hz, 3H).

541 **(*E*)-2-hydroxy-3-methyldec-4-enoic acid (bemeth#2, 12)**



542

543 To a solution of **15** (76 mg, 0.33 mmol) in 1,4-dioxane (2 mL), lithium hydroxide monohydrate
544 (42 mg, 1.00 mmol) in H_2O (1 mL) was added, and the resulting mixture was stirred at 60 °C for
545 12 hours. Glacial acetic acid (0.5 mL) was added, and the reaction was concentrated *in vacuo*.
546 Flash column chromatography on silica using a gradient of 0-50% ethyl acetate in hexanes
547 afforded **12** (46.7 mg, 70%) as a clear oil.

548 For ^{1}H and ^{13}C NMR spectroscopic data of the major diastereomer of bemeth#2 (**12)**, see
549 **Supplementary Table S3**.

550 Minor diastereomer, ^{1}H NMR (600 MHz, methanol-d_4): δ (ppm) δ (ppm) 5.56-5.41 (m, 2H), 3.96
551 (d, *J* = 4.6 Hz, 1H), 2.54 (m, 1H), 2.00 (m, 2H), 1.43-1.25 (m, 6H), 1.00 (d, *J* = 6.9 Hz, 3H), 0.90
552 (t, *J* = 7.0 Hz, 3H).

553

554

555

556 **Data availability.** All data analyzed during this study are included in the manuscript and
557 supporting files. MS/MS data is available via MassIVE under accession number:
558 MSV000087885. Helf MJ, Fox BW, Artyukhin AB, Zhang YK, Schroeder FC (2021) MassIVE ID
559 MSV000087885. Comparative metabolomics with Metaboseek reveals functions of peroxisomal
560 α-oxidation in *C. elegans* [https://doi:10.25345/C5P24X]. Source data are provided with this
561 paper.

562 **Code availability.** Metaboseek is available as an R package, with installation instructions for
563 Windows, macOS, and Linux. A preconfigured R-portable installation is available as installer or
564 .zip file for Windows. All source code and downloads are available at
565 https://github.com/mjhelf/Metaboseek or https://doi.org/10.5281/zenodo.3360087. Code
566 documentation and a tutorial vignette are part of the Metaboseek R package[58]. Individual
567 functions for spectra comparison, merging spectra and filtering molecular formulae have been
568 moved to the companion R package MassTools[59], with source code and documentation
569 available at https://github.com/mjhelf/MassTools and https:/doi.org/10.5281/zenodo.5725620.
570 The Metaboseek tutorial is also available at https://metaboseek.com/doc.

571 **Acknowledgements.** This research was supported in part by the National Institutes of Health
572 (R35GM131877 and U01GM110714 to F.C.S.). F.C.S. is a Faculty Scholar of the Howard
573 Hughes Medical Institute. M.J.H received a Research Fellowship from the Deutsche
574 Forschungsgemeinschaft (DFG), Project Number 386228702. Some strains used in this work
575 were provided by the CGC, which is funded by the NIH Office of Research Infrastructure
576 Programs (P40 OD010440). *B0334.3(tm6725)* was obtained from the National Bioresource
577 Project, Tokyo, Japan. We thank Diana Carolina Fajardo Palomino and Gary Horvath for
578 technical support and David Kiemle for assistance with NMR spectroscopy. We thank Dr. Yong-
579 Uk Lee and the Walhout lab for kindly sharing the JW1653-1 *E. coli* strain.

580 **Author Contributions.** FCS supervised the study. MJH and FCS conceived the Metaboseek
581 platform. MJH developed the Metaboseek platform. BWF and ABA performed chemical and
582 biological experiments. YKZ performed syntheses. BWF, MJH, and FCS wrote the paper with
583 input from all authors.

584 **Competing Interests.** The authors declare no competing interests.

585

586 **References**

587 1. Zamboni, N., Saghatelian, A. & Patti, G. J. Defining the Metabolome: Size, Flux, and
588 Regulation. *Molecular Cell* vol. 58 699–706 (2015).

589 2. Joshua, W. *et al.* A novel 3-hydroxysteroid dehydrogenase that regulates reproductive
590 development and longevity. *PLoS Biol.* (2012).

591 3. Watson, E. *et al.* Interspecies systems biology uncovers metabolites affecting C. elegans
592 gene expression and life history traits. *Cell* (2014) doi:10.1016/j.cell.2014.01.047.

593 4. O'Donnell, M. P., Fox, B. W., Chao, P. H., Schroeder, F. C. & Sengupta, P. A
594 neurotransmitter produced by gut bacteria modulates host sensory behaviour. *Nature*
595 (2020) doi:10.1038/s41586-020-2395-5.

596 5. Funabashi, M. *et al.* A metabolic pathway for bile acid dehydroxylation by the gut
597 microbiome. *Nature* **582**, 566–570 (2020).

598 6. Patti, G. J., Yanes, O. & Siuzdak, G. Innovation: Metabolomics: the apogee of the omics
599 trilogy. *Nature Reviews Molecular Cell Biology* (2012) doi:10.1038/nrm3314.

600 7. Guijas, C., Montenegro-Burke, J. R., Warth, B., Spilker, M. E. & Siuzdak, G.
601 Metabolomics activity screening for identifying metabolites that modulate phenotype. *Nat.*
602 *Biotechnol. 2018 364* **36**, 316–320 (2018).

603 8. Wishart, D. S. Metabolomics for investigating physiological and pathophysiological
604 processes. *Physiol. Rev.* **99**, 1819–1875 (2019).

605 9. Johnson, C. H., Ivanisevic, J. & Siuzdak, G. Metabolomics: Beyond biomarkers and
606 towards mechanisms. *Nature Reviews Molecular Cell Biology* (2016)
607 doi:10.1038/nrm.2016.25.

608 10. Walker, D. I. *et al.* The metabolome: A key measure for exposome research in
609 epidemiology. *Curr. Epidemiol. reports* **6**, 93 (2019).

610 11. Da Silva, R. R., Dorrestein, P. C. & Quinn, R. A. Illuminating the dark matter in
611 metabolomics. *Proc. Natl. Acad. Sci. U. S. A.* **112**, 12549–12550 (2015).

612 12. Sindelar, M. & Patti, G. J. Chemical Discovery in the Era of Metabolomics. *J. Am. Chem.*
613 *Soc.* **142**, 9097–9105 (2020).

614 13. Mahieu, N. G. & Patti, G. J. Systems-Level Annotation of a Metabolomics Data Set
615 Reduces 25 000 Features to Fewer than 1000 Unique Metabolites. *Anal. Chem.* **89**,
616 10397–10406 (2017).

617 14. Baker, M. Metabolomics: From small molecules to big ideas. *Nat. Methods* (2011)
618 doi:10.1038/nmeth0211-117.

619 15. Wanders, R. J. A., Waterham, H. R. & Ferdinandusse, S. Metabolic interplay between
620 peroxisomes and other subcellular organelles including mitochondria and the
621 endoplasmic reticulum. *Frontiers in Cell and Developmental Biology* (2016)
622 doi:10.3389/fcell.2015.00083.

623 16. Artyukhin, A. B. *et al.* Metabolomic 'dark Matter' Dependent on Peroxisomal β-Oxidation
624 in Caenorhabditis elegans. *J. Am. Chem. Soc.* (2018) doi:10.1021/jacs.7b11811.

625 17. Pluskal, T., Castillo, S., Villar-Briones, A. & Orešič, M. MZmine 2: Modular framework for
626 processing, visualizing, and analyzing mass spectrometry-based molecular profile data.
627 *BMC Bioinformatics* **11**, (2010).

628 18. Tsugawa, H. *et al.* MS-DIAL: Data-independent MS/MS deconvolution for comprehensive
629 metabolome analysis. *Nat. Methods* **12**, 523–526 (2015).

630 19. Tsugawa, H. *et al.* A lipidome atlas in MS-DIAL 4. *Nat. Biotechnol. 2020 3810* **38**, 1159–
631 1163 (2020).

632 20. Tautenhahn, R., Patti, G. J., Rinehart, D. & Siuzdak, G. XCMS online: A web-based
633 platform to process untargeted metabolomic data. *Anal. Chem.* **84**, 5035–5039 (2012).

634 21. Chong, J. *et al.* MetaboAnalyst 4.0: Towards more transparent and integrative
635 metabolomics analysis. *Nucleic Acids Res.* **46**, W486–W494 (2018).

636 22. Kuhl, C., Tautenhahn, R., Böttcher, C., Larson, T. R. & Neumann, S. CAMERA: An
637 Integrated Strategy for Compound Spectra Extraction and Annotation of Liquid
638 Chromatography/Mass Spectrometry Data Sets. *Anal. Chem.* **84**, 283–289 (2012).

639 23. Dührkop, K. *et al.* SIRIUS 4: a rapid tool for turning tandem mass spectra into metabolite
640 structure information. *Nat. Methods* (2019) doi:10.1038/s41592-019-0344-8.

641 24. Djoumbou-Feunang, Y. *et al.* Cfm-id 3.0: Significantly improved esi-ms/ms prediction and

642 compound identification. *Metabolites* **9**, (2019).

643 25. Watrous, J. *et al.* Mass spectral molecular networking of living microbial colonies. *Proc.*
644 *Natl. Acad. Sci. U. S. A.* **109**, E1743–E1752 (2012).

645 26. Duy Nguyen, D. *et al.* MS/MS networking guided analysis of molecule and gene cluster
646 families. (2013) doi:10.1073/pnas.1303471110.

647 27. Nothias, L. F. *et al.* Feature-based molecular networking in the GNPS analysis
648 environment. *Nat. Methods* **17**, 905–908 (2020).

649 28. Foulon, V. *et al.* Purification, molecular cloning, and expression of 2-hydroxyphytanoyl-
650 CoA lyase, a peroxisomal thiamine pyrophosphate-dependent enzyme that catalyzes the
651 carbon-carbon bond cleavage during alpha -oxidation of 3-methyl-branched fatty acids.
652 *Proc. Natl. Acad. Sci.* **96**, 10039–10044 (1999).

653 29. Mukherji, M. *et al.* The chemical biology of branched-chain lipid metabolism. *Progress in*
654 *Lipid Research* (2003) doi:10.1016/S0163-7827(03)00016-X.

655 30. Croes, K., Foulon, V., Casteels, M., Van Veldhoven, P. P. & Mannaerts, G. P. Phytanoyl-
656 CoA hydroxylase: Recognition of 3-methyl-branched acyl-CoAs and requirement for GTP
657 or ATP and Mg2+ in addition to its known hydroxylation cofactors. *J. Lipid Res.* **41**, 629–
658 636 (2000).

659 31. Wanders, R. J. A., Komen, J. & Ferdinandusse, S. Phytanic acid metabolism in health
660 and disease. *Biochimica et Biophysica Acta - Molecular and Cell Biology of Lipids* (2011)
661 doi:10.1016/j.bbalip.2011.06.006.

662 32. Bandu, M. L., Grubbs, T., Kater, M. & Desaire, H. Collision induced dissociation of alpha
663 hydroxy acids: Evidence of an ion-neutral complex intermediate. *Int. J. Mass Spectrom.*
664 (2006) doi:10.1016/j.ijms.2006.01.004.

665 33. Wanders, R. J. A., Komen, J. & Kemp, S. Fatty acid omega-oxidation as a rescue
666 pathway for fatty acid oxidation disorders in humans. *FEBS Journal* (2011)
667 doi:10.1111/j.1742-4658.2010.07947.x.

668 34. Gao, A. W. *et al.* A sensitive mass spectrometry platform identifies metabolic changes of
669 life history traits in C. elegans. *Sci. Rep.* **7**, 1–14 (2017).

670 35. Artyukhin, A. B., Yim, J. J., Cheong Cheong, M. & Avery, L. Starvation-induced collective
671 behavior in C. elegans. *Sci. Rep.* (2015) doi:10.1038/srep10647.

672 36. Watts, J. L. & Ristow, M. Lipid and carbohydrate metabolism in Caenorhabditis elegans.
673 *Genetics* **207**, 413–446 (2017).

674 37. Kniazeva, M., Crawford, Q. T., Seiber, M., Wang, C. Y. & Han, M. Monomethyl branched-
675 chain fatty acids play an essential role in Caenorhabditis elegans development. *PLoS*
676 *Biol.* **2**, (2004).

677 38. Grogan, D. W. & Cronan, J. E. Cyclopropane ring formation in membrane lipids of
678 bacteria. *Microbiol. Mol. Biol. Rev.* **61**, 429–441 (1997).

679 39. Brooks, K. K., Liang, B. & Watts, J. L. The influence of bacterial diet on fat storage in C.
680 elegans. *PLoS One* (2009) doi:10.1371/journal.pone.0007545.

681 40. Baba, T. *et al.* Construction of Escherichia coli K-12 in-frame, single-gene knockout
682 mutants: The Keio collection. *Mol. Syst. Biol.* (2006) doi:10.1038/msb4100050.

683 41. Kaul, T. K., Reis Rodrigues, P., Ogungbe, I. V., Kapahi, P. & Gill, M. S. Bacterial fatty
684 acids enhance recovery from the dauer larva in Caenorhabditis elegans. *PLoS One*
685 (2014) doi:10.1371/journal.pone.0086979.

686 42. Caligiani, A., Nocetti, M., Lolli, V., Marseglia, A. & Palla, G. Development of a
687 Quantitative GC-MS Method for the Detection of Cyclopropane Fatty Acids in Cheese as
688 New Molecular Markers for Parmigiano Reggiano Authentication. *J. Agric. Food Chem.*
689 (2016) doi:10.1021/acs.jafc.6b00913.

690 43. Lolli, V., Dall'Asta, M., del Rio, D. & Caligiani, A. Identification of Cyclopropane Fatty
691 Acids in Human Plasma after Controlled Dietary Intake of Specific Foods. *Nutrients* **12**,
692 1–10 (2020).

693 44. Sud, M. *et al.* Metabolomics Workbench: An international repository for metabolomics
694 data and metadata, metabolite standards, protocols, tutorials and training, and analysis
695 tools. *Nucleic Acids Res.* **44**, D463–D470 (2016).

696 45. Aron, A. T. *et al.* Reproducible molecular networking of untargeted mass spectrometry
697 data using GNPS. *Nat. Protoc.* **15**, 1954–1991 (2020).

698 46. Alseekh, S. *et al.* Mass spectrometry-based metabolomics: a guide for annotation,
699 quantification and best reporting practices. *Nat. Methods* **18**, 747–756 (2021).

700 47. Wang, M. *et al.* Sharing and community curation of mass spectrometry data with Global
701 Natural Products Social Molecular Networking. *Nature Biotechnology* vol. 34 828–837
702 (2016).

703 48. Petras, D. *et al.* GNPS Dashboard: Collaborative Analysis of Mass Spectrometry Data in
704 the Web Browser. *bioRxiv* 2021.04.05.438475 (2021) doi:10.1101/2021.04.05.438475.

705 49. Horai, H. *et al.* MassBank: a public repository for sharing mass spectral data for life
706 sciences. *J. Mass Spectrom.* **45**, 703–714 (2010).

707 50. Blaženović, I., Kind, T., Ji, J. & Fiehn, O. Software Tools and Approaches for Compound
708 Identification of LC-MS/MS Data in Metabolomics. *Metabolites* **8**, (2018).

709 51. Schrimpe-Rutledge, A. C., Codreanu, S. G., Sherrod, S. D. & McLean, J. A. Untargeted
710 Metabolomics Strategies—Challenges and Emerging Directions. *J. Am. Soc. Mass*
711 *Spectrom.* **27**, 1897–1905 (2016).

712 52. Viant, M. R., Kurland, I. J., Jones, M. R. & Dunn, W. B. How close are we to complete
713 annotation of metabolomes? *Current Opinion in Chemical Biology* vol. 36 64–69 (2017).

714 53. Mize, C. E., Steinberg, D., Avigan, J. & Fales, H. M. A pathway for oxidative degradation
715 of phytanic acid in mammals. *Biochem. Biophys. Res. Commun.* **25**, 359–365 (1966).

716 54. Mize, C. E. *et al.* A major pathway for the mammalian oxidative degradation of phytanic
717 acid. *Biochim. Biophys. Acta - Lipids Lipid Metab.* **176**, 720–739 (1969).

718 55. Barstead, R. *et al.* Large-scale screening for targeted knockouts in the caenorhabditis
719 elegans genome. *G3 Genes, Genomes, Genet.* **2**, 1415–1425 (2012).

720 56. Artyukhin, A. B., Schroeder, F. C. & Avery, L. Density dependence in Caenorhabditis
721 larval starvation. *Sci. Rep.* (2013) doi:10.1038/srep02777.

722 57. MC, C. *et al.* A cross-platform toolkit for mass spectrometry and proteomics. *Nat.*
723 *Biotechnol.* **30**, 918–920 (2012).

724 58. Helf, M. J. mjhelf/Metaboseek: v0.9.9. (2021) doi:10.5281/ZENODO.3360087.

59. Helf, M. J. mjhelf/MassTools: MassTools Version 0.2.12. (2021) doi:10.5281/ZENODO.5725620.

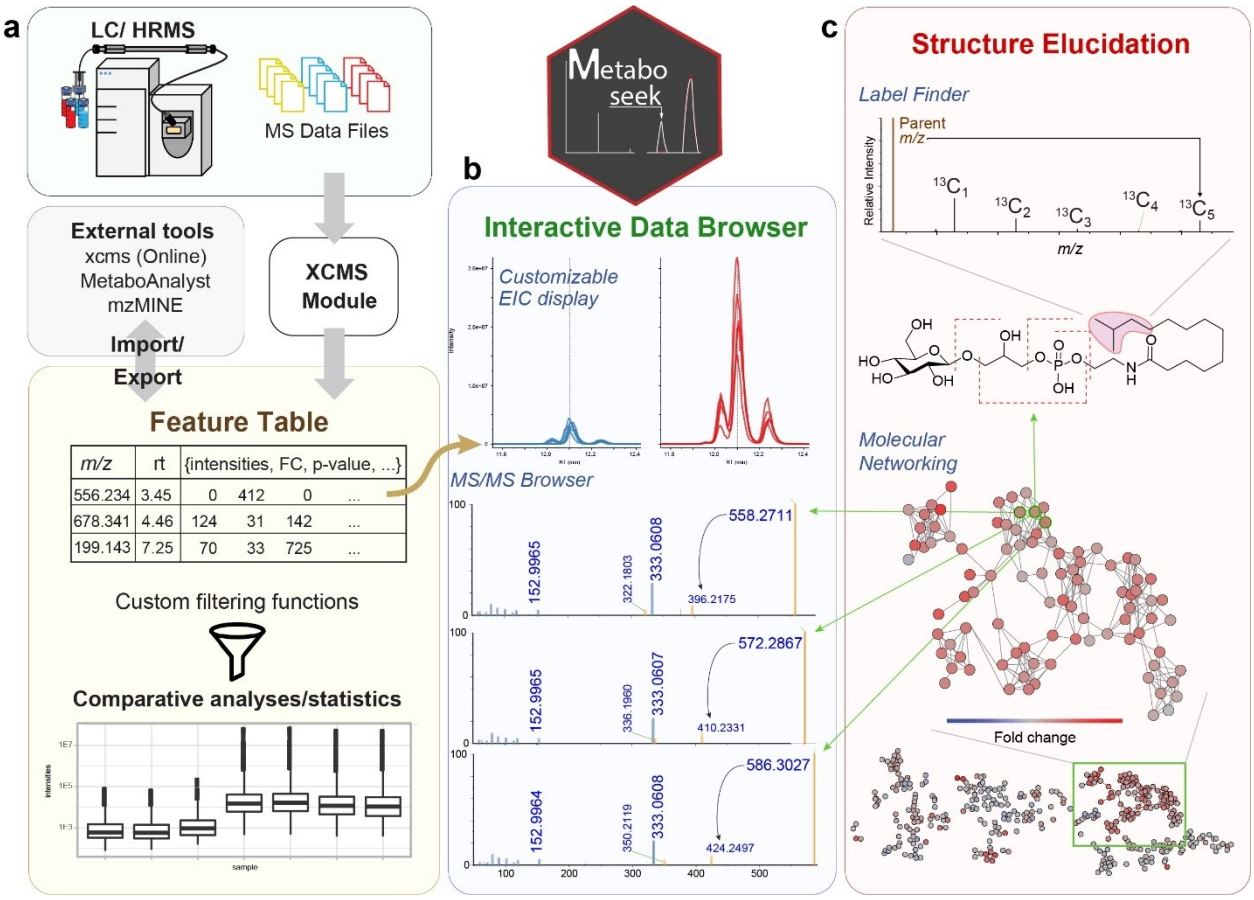


Figure 1. Comparative Metabolomics with Metaboseek. **a,** Metaboseek includes an integrated XCMS module for feature detection and feature grouping (with CAMERA annotation) and accepts feature tables generated by other software. **b,** Features can be annotated and prioritized using extensive filtering options and integrated statistics tools. Raw data for each molecular feature can be browsed rapidly, including associated EICs, MS1 and MS/MS spectra. **c**, The data browser interacts with a suite of structure elucidation tools, e.g., SIRIUS-based molecular formula and structure prediction, the *Label Finder* to identify isotope-labeled compounds, and the MS/MS pattern finder to identify MS features with characteristic fragmentation patterns.

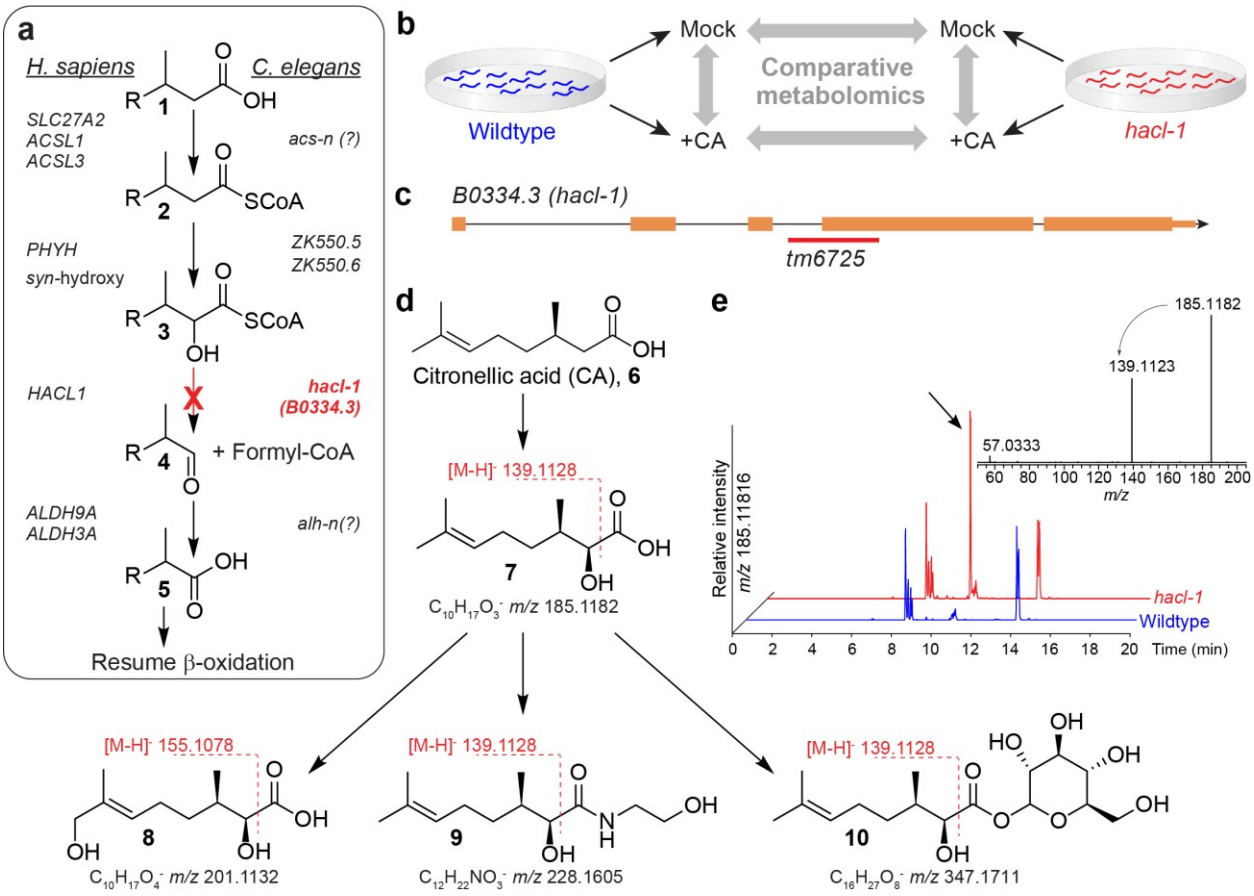


Figure 2. Comparative metabolomics of CA-fed *hacl-1* mutants. **a,** Conservation of peroxisomal α-oxidation in *C. elegans* and humans. **b,** CA-feeding experiment. *C. elegans* larvae were supplemented with CA, (**6**) followed by comparative analysis with Metaboseek. **c,** *hacl-1(tm6725)* mutants harbor a genomic deletion (red line) spanning the indicated exon splice junction, orange rectangles represent exons, black lines represent introns. **d,** Representative shunt metabolites accumulating in *hacl-1* following CA supplementation. The majority of shunt metabolites fragment between the carbonyl- and α-carbon during MS/MS, which is characteristic of α-hydroxy fatty acids. **e,** HPLC-MS (negative ion) EIC for *hacl-1*-enriched feature with *m/z* 185.1182, corresponding to **7** (arrow). Its MS/MS spectrum reveals a product ion at *m/z* 139.1123, corresponding to neutral loss of formic acid.

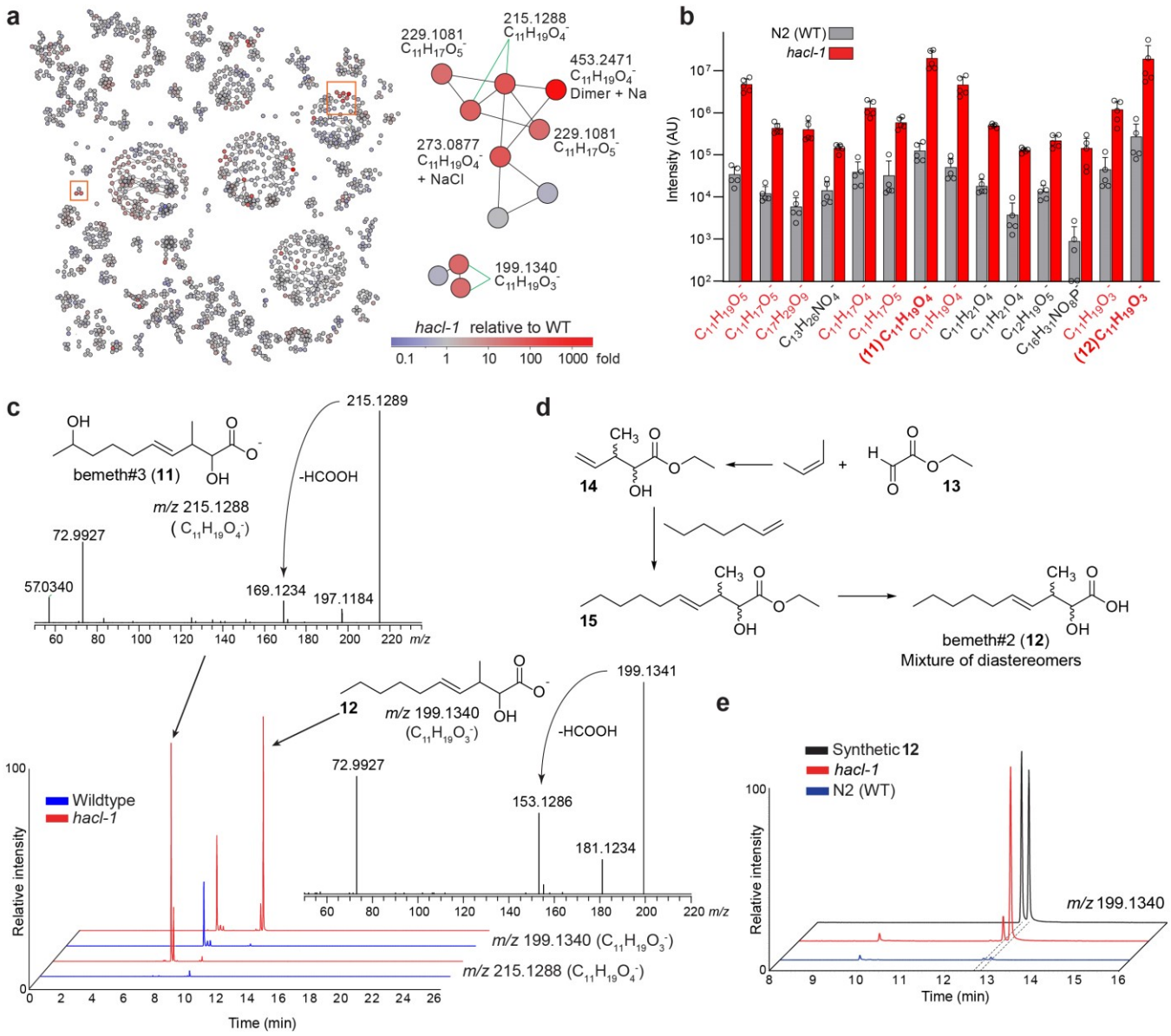


Figure 3. Endogenous metabolites accumulating in *hacl-1* mutant larvae. **a,** MS/MS network highlighting differential abundance in *hacl-1* mutants relative to WT. Subnetworks of interest (orange boxes) are shown enlarged. **b,** Quantification of endogenous metabolites ten-fold or more enriched in *hacl-1* larvae relative to WT at $p < 0.05$. Metabolites in red produce an MS/MS product ion with *m/z* 72.993. Data represent five independent experiments and bars means ± standard deviation. **c,** Representative HPLC-MS (negative ion) EICs and MS/MS spectra for most abundant differential metabolites in **b**, bemeth#3 (**11**) and bemeth#2 (**12**). **d,** Overview of synthetic scheme to afford bemeth#2 (see Methods for details). **e,** Comparison of HPLC-MS (negative ion) EICs for synthetic diastereomers of bemeth#2 (**12**) and the corresponding metabolites in *exo*-metabolome extracts from WT and *hacl-1* larvae.

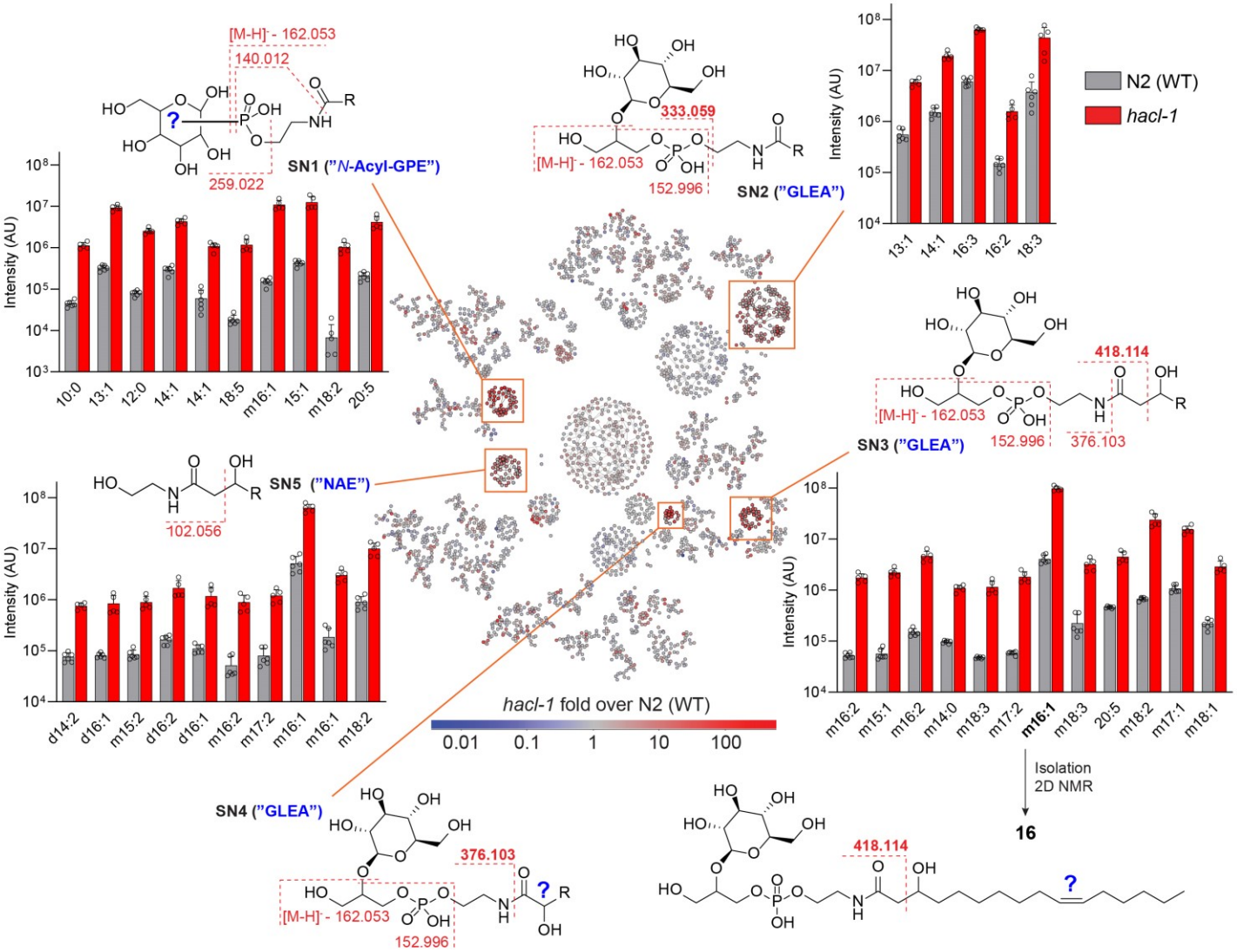


Figure 4. MS/MS network comparing *exo*-metabolomes of *hacl-1* and WT adults. Proposed structures and major fragmentation reactions are shown for five subnetworks (SN1 – SN5, orange boxes). Example compounds in the bar graphs are at least 10-fold enriched in *hacl-1* mutants relative to WT and satisfy mean intensity criteria (10^6 for SN1, SN2, SN3; $5*10^5$ for SN5). R represents an acyl group with N carbons and n degrees of unsaturation (N:n), preceded by *m* or *d* for *m*ono- or *d*i-oxygenated. The structure of the most abundant metabolite in SN3, GLEA-m16:1 (**16**), was characterized via 2D NMR spectroscopy. For compounds from SN4, see **Supplementary Figure 10**. Data represent six (WT) or five (*hacl-1*) samples from three biologically independent experiments and bars means ± standard deviation.

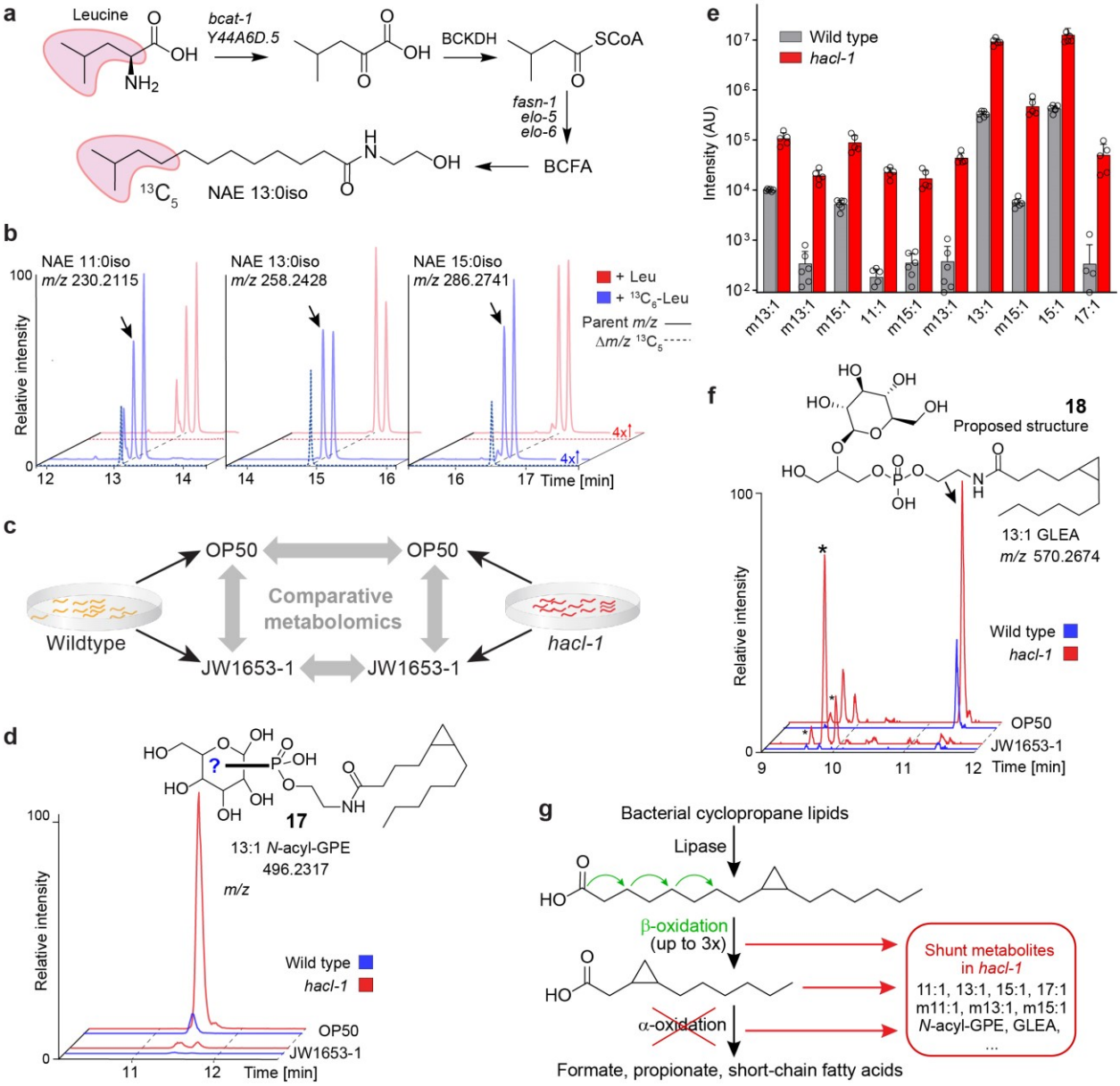


Figure 5. Cyclopropane-containing glycolipids are enriched in *hacl-1* mutants. **a,** Leucine metabolism feeds into branched-chain fatty acid (BCFA) biosynthesis in *C. elegans*. **b,** Pairwise analysis of Leu or $^{13}C_6$-Leu-supplemented worms using the *Label Finder* revealed $^{13}C_5$-enriched NAEs derived from BCFAs. Shown are representative HPLC-MS (ESI+) EICs and dotted lines represent incorporation of $^{13}C_5$ ($\Delta m/z$, 5.0167), as visualized using Metaboseek *Mass Shifts.* **c,** Study design for comparative metabolomics of WT and *hacl-1* worms fed either *E. coli* OP50 or cyclopropane fatty acid-deficient *E. coli* JW1653-1. **d**, Representative HPLC-MS (ESI-) EIC for 496.2317, corresponding to cyclopropane-containing *N*-acyl GPE 13:1 (**17**). Shown structure was proposed based on MS/MS fragmentation and absence in JW1653-1. **e**, Quantification of

779 *N*-acyl GPEs that were absent from worms fed JW1653-1. Data represent six (WT) or five (*hacl-*
780 *1*) samples from three biologically independent experiments and bars means ± SD. **f**,
781 Representative HPLC-MS (ESI-) EIC for *m/z* 570.2674, corresponding to cyclopropane-
782 containing GLEA 13:1 (**18**). Shown structure was proposed based on MS/MS fragmentation and
783 absence in JW1653-1. Asterisks (*) mark isobaric *hacl-1*-enriched features that are not
784 impacted by JW1653-1 diet. **g,** Proposed metabolism of bacterial 17:1 (or 19:1) cyclopropane
785 lipids. Three (or four) rounds of β-oxidation would produce an 11:1 cyclopropane fatty acid
786 unsuitable for further β-oxidation that could be a substrate for α- or ω-oxidation.

787