Genetic basis of cuticular hydrocarbon variation in the desert ant

Shani Inbar, Besan Saied, Pnina Cohen, Zeev Frenkel, Yoann Pellen, Abraham B. Korol, Eyal Privman[#]
Department of Evolutionary and Environmental Biology, Institute of Evolution, University of Haifa, Haifa, Israel
[#] Corresponding author: eprivman@univ.haifa.ac.il

Abstract

Cuticluar hydrocarbons (CHCs) are a ubiquitous component of insect cuticles that are used for a wide range of chemical signaling functions, especially recognition. Recognition and other signals are vital for the maintainance of insularity and cooperation in social insect colonies. Therefore, we expect natural selection on the composition and variabitlity of social insect CHC profiles. Selection on these signals may result in the evolution of genetic polymorphism affecting variations in CHC profiles. Here we tested for a genetic basis of CHC variation in the desert ant *Cataglyphis niger*. We applied a genomic mapping appraoch to a cohort of brothers from the same nest to reduce noise from environmental effects and achieve a clear statistical signal for association between CHCs and quantitative trait loci (QTL). This analyses identified 19 QTLs associated with 8 our of 31 tested CHCs, and one QTL associated with total CHC quantity. These QTLs are located on 11 different chromosomes, including some where QTLs of different CHCs overlap. Each QTL explains between 13-25% of the CHC variation. We highlight several candidate genes in the QTLs identified, including fatty acid elongase and reductase genes. Our results reveal a polygenic genomic architecture underlying CHC variation in ants and open new research avenues into the genetic basis and evolution of chemical signaling in social insects.

Introduction

Nestmate recognition is a fundamental trait of social insects. The ability to discriminate nestmates from foreign intruders is essential for colony integrity and protection of colony resources. Long-chain cuticular hydrocarbons (CHCs) are the main chemical cues underlying nestmate recognition (Holldobler and Michener 1980; Singer 1998). In this study, we test whether CHC variability has a genetic basis and identified candidate genes that may affect this important phenotype. We used the desert ant *Cataglyphis niger* as our study organism, in which nestmate recognition was extensively studied using chemical-behavioral approaches.

C. niger is one of the first species where CHCs were empirically demonstrated to be major components of the nestmate recognition cues (Lahav, et al. 1999).

CHCs are the main lipid constituents of the insect cuticle (reviewed in Blomquist and Bagneres 2010; Blomquist, et al. 2020). The ancestral role of CHCs was probably to provide a waterproof layer preventing desiccation. Subsequently, they evolved a secondary role as chemical cues for signaling, communication, and recognition. Insect CHCs have variable chain lengths (20 to 50 carbons) and variable number and position of double bonds and methyl branches, thus allowing a great diversity that affords high information content as chemical cues. CHCs function as sex pheromones in many insects, as attractants, as fertility signals, and as chemical cues for species and sub-species recognition that facilitates mate choice (Blomquist and Bagneres 2010). Different social insect castes generally display distinct CHC profiles, including distinction between fertile individuals (queens, gamergates, or egg laying workers) and their less fertile or non-reproductive nestmates (Howard, et al. 1982; Nunes, et al. 2009; Kleeberg, et al. 2017; Derstine, et al. 2018; Sprenger and Menzel 2020). Signaling by CHCs have been extensively studied in social insects, where they make up an important component of the elaborate communication system necessary for coordinating complex collective behaviors (Hefetz 2007).

The enzymatic pathways responsible for the biosynthesis of CHCs were not characterized in social insects, however, a large body of literature describes such enzymes in model insect species, most notably *Drosophila melanogaster* (Blomquist and Bagneres 2010; Holze, et al. 2021). CHCs are *de novo* synthesized from two- or three-carbon units. Short saturated fatty acids (up to 18 carbons) are first synthesized by fatty acid synthases. These precursors are then elaborated by a series of enzymes that elongate the carbon chain (elongases), introduce double bonds (desaturases), and the final enzymatic steps that convert the fatty acid to a hydrocarbon, including reductases and decarboxylases. Homologs of these enzymes are found in ant genomes, and are the natural candidates for the genetic basis of ant CHCs. Beyond the genes involved in synthesis of CHCs, other genes may be involved in regulation of the quantitates of CHCs produced and secreted through the cuticle. These genes may function in any stage of the long road from the site of CHC production to the external surface of the cuticle. CHCs can be either synthesized de novo or from diet-derived fatty acids (Liang and Silverman 2000; Blomquist and Bagneres 2010). They are thought to be produced primarily in the fat body, in cells known as oenocytes (Blomquist and Bagneres 2010).

Following their initial synthesis, CHCs are transported through the hemolymph and eventually secreted and deposited on the cuticle.

One of the best understood social functions of CHCs is in nestmate recognition, a fundamental and vital function in social insect colonies. Nestmate recognition is primarily based on chemical recognition, which was suggested to involve matching an olfactory perception of chemical cues to an experience-derived neural template (Ozaki and Hefetz 2014). Different species display different mixtures of CHCs (Martin and Drijfhout 2009a; d'Ettorre and Lenoir 2010), which suggests rapid evolution, possibly driven by strong selection. Intra-specific quantitative variation generates colony-specific CHC profiles, which are used for nestmate recognition. Experimental manipulation demonstrated the role of CHCs in nestmate recognition in multiple ant species (Bonavita-Cougourdan, et al. 1987; Nowbahari, et al. 1990; Akino, et al. 2004; Ozaki, et al. 2005; Martin, et al. 2008), including in our study species *Cataglyphis niger* (Lahav, et al. 1999). These studies transferred the hydrocarbon fraction of extracts from one ant to another, which caused ants to attack nestmates or to reduce aggression towards non-nestmates. Some studies went a step further: they synthesized specific hydrocarbons and tested their effect on nestmate recognition (Akino, et al. 2004; Martin, et al. 2008).

Recognition cues may be heritable and/or environmentally derived. Heritability of CHC profiles would imply that quantitative variations in CHCs are partially explained by genetic polymorphism. Both genetic and environmental effects were documented in *Drosophila*, and some of the underlying genes were identified (reviewed in Ferveur 2005). Multiple studies suggest that quantitative variation in CHCs is heritable in diverse insect species including crickets (Thomas and Simmons 2008), termites (Dronnet, et al. 2006), and parasitoid wasps (Niehuis, et al. 2011). Empirical studies in social insects suggest that CHC variation is affected by both environmental and genetic factors. For example, a strong effect of the diet was reported in the Argentine ant (Liang and Silverman 2000). Pathogens may also have an effect on nestmate recognition in ants (Tranter, et al. 2015; Leclerc and Detrain 2016). A study in honeybees reported a pathogen's effect on CHCs and nestmate recognition (Cappa, et al. 2019). Other potential factors may include abiotic effects such as temperature and humidity (Silva, et al. 2024). Nevertheless, multiple studies suggested that ant CHCs are heritable to a certain degree. Several studies demonstrated under controlled conditions and uniform diet that intercolonial variation in CHC profiles is retained as well as the ability for

nestmate recognition (Nowbahari, et al. 1990; Lahav, et al. 1999; Akino, et al. 2004; Ozaki, et al. 2005; van Zweden, et al. 2009). Behavioral experiments on different populations of the ant *Leptothorax longispinosus* revealed an association between aggression and genetic dissimilarity (Stuart and Herbers 2000). Isolation of *Formica exsecta* workers revealed a weak patriline effect on CHCs, indicating that different alleles inherited from different fathers had an effect on the CHC profiles of their daughters (Martin, et al. 2012). A study in the ant *Monomorium pharaonis* tested for heritability of CHCs by systematic intercrossing eight colonies over nine generations (Walsh, et al. 2020), resulting in low to moderate estimated heritability: narrow-sense heritability h^2 between 0 and 0.36. These studies provide evidence for heritability of CHCs in diverse ant species.

The vital function of nestmate recognition is expected to have large fitness consequences, including "red queen" dynamics due to selection against social parasites (Blomquist and Bagneres 2010). This hypothesis is supported by the great interspecific diversity of CHCs and their rapid evolution (Martin and Drijfhout 2009a; d'Ettorre and Lenoir 2010). Some studies found stronger differentiation between colonies by methyl-branched vs. straight alkanes (e.g., Dahbi, et al. 1996). The abundance of branched alkanes and unsaturated alkenes in ant CHCs, which provide greater information content as cues while compromising protection from desiccation, was suggested to be the result of selection for signaling functions (d'Ettorre and Lenoir 2010; Sturgis and Gordon 2012). The exceptional complexity of ant CHC profiles suggests selection for recognition accuracy. If genetic polymorphism arises such that it affects nestmate recognition cues, selection for distinctness of colony profiles is expected to increase the frequency of novel alleles. The requirement for distinctness of colony CHC profiles should lead to selection in favor of rare alleles, i.e. frequency-dependent selection resulting in balancing selection (Ayala and Campbell 1974; Jongepier and Foitzik 2016). Such processes result in elevated polymorphism and the maintenance of highly diverged alleles in the population over long evolutionary time, as observed for example in mammalian Major Histocompatibility Complex (MHC) genes (Hughes and Nei 1988) and the honeybee Complementary Sex Determination (CSD) locus (Cho, et al. 2006). Such evolutionary dynamics are expected to result in the accumulation of polymorphic loci that affect the display of ant CHCs quantitatively, i.e., quantitative trait loci (QTLs).

Here, we conducted a QTL mapping study to test whether quantitative variation of CHCs is associated with genetic polymorphism in the desert ant *Cataglyphis niger*. We produce the

genomic infrastructures necessary for mapping associated genomic loci – a chromosome-level reference genome sequence. We conduct reduced representation genomic sequencing of a cohort of brothers to identify polymorphic loci associated with CHC variation (QTLs) and characterize the genomic architecture of these traits. The use of haploid males affords valuable advantages in terms of the simplicity of the underlying genetics and minimal environmental variation (see Methods). These QTLs identified in males are likely to also have effects on CHCs in workers and queens because many of the same CHCs are observed in all castes of the species (Hefetz 2007). We highlight promising candidate genes in these loci. The discovery of genetic loci associated with CHC variation serves as a first step in the study of the molecular mechanisms underlying nestmate recognition and their evolution in a social insect species.

Methods

Samples

Our analysis was based on a sample of 117 brothers, collected on 3/5/2016 from a single *C. niger* nest in Betzet Beach in the northern Israeli coastline (colony BZT50, 33°03'32N 35°06'15E) and frozen in -80°C. This colony belongs to a population that was previously referred to as *Cataglyphis drusus*[39], but our species delimitation analysis showed that it is part of the species described as *C. niger* that is widespread across Israel. Our population structure analysis showed that all populations across Israel are not differentiated by their nuclear genomic DNA[40]. We note that in some areas in Israel *C. niger* colonies have multiple queens (polygyne)[40] and that our recent study revealed that this social polymorphism is associated with a supergene on chromosome 2 (Lajmi, et al. 2024). A supergene is a large non-recombining chromosomal region, analogous to sex chromosomes. However, the study population used here is purely monogyne (each colony is headed by a single queen)[39,40]. The supergene is always homozygous (MM) in monogyne colonies (the P haplotypes is only found in polygyne colonies), so we do not expect it to affect the QTL mapping analysis in this study.

Using a cohort of brothers for QTL mapping affords several advantages. Ants have haplodiploid sex determination, where haploid males develop from unfertilized eggs, and diploid females derive from fertilized eggs. Thus, males have no fathers. They only receive one of their mother's alleles in each locus. And so, a cohort of brothers resents the products

172 of maternal meiosis. Such a cohort servses as a simple mapping population similar to a back-
173 cross experiment that is traditionally used for QTL mapping studies. Moreover, recessive
174 alleles are more readily revealed in haploid progeny because samples are hemizygous at all
175 loci. For the same reason, similar studies using haploid males were conducted in another
176 hymenopteran species – the parasitoid wasp *Nasonia vitripennis* (Niehuis, et al. 2011). This
177 approach successfully identified a large number of QTLs associated with many of the CHCs
178 analyzed. Another advantage is that CHC variation among brothers sampled from the same
179 nest is likely to be minimally affected by environmental factors because they all experienced
180 a similar environment in the nest. For example, we can safely assume they were fed the same
181 diet and exposed to the same microbes by the workers that cared for them. Males are also
182 expected to be more uniform than workers because they do not have distinct task groups that
183 are exposed to different environments (e.g. nurses vs. foragers).
184
185 **Chemical analysis**
186 Samples were extracted in 100 µl hexane, containing 200 ng tetracosane (C24) as internal
187 standard. Compound identification was done by combined GC/MS and was based on their
188 fragmentation patterns, as compared to authentic standards. Sample quantification was done
189 by gas chromatography, using a VF-5ms capillary column that was temperature-programmed
190 from 150°C (with 1 min initial hold) to 300°C at a rate of 5°C per min, with a final hold of 15
191 min. Peak integration was performed using the program Galaxie Varian 1.9. Absolute
192 quantities were calculated relative to the internal standard.
193
194 **Reference genome sequencing, assembly and annotation**
195 Identification of the genes inside QTLs requires a genome assembly of sufficient contiguity,
196 that is, genomic scaffolds should be large enough so that the markers of the linkage map can
197 be linked to the genes within the confidence interval for the position of a QTL. We already
198 had a reference genome assembly for *C. niger* that was assembled from Illumina short-read
199 sequencing of a single haploid male (version Cnig_gn1; Yahav and Privman 2019). However,
200 this assembly was rather fragmented, with a scaffold N50 size of only 17.9 Kb. Therefore, we
201 conducted Pacbio long-read sequencing to improve the assembly contiguity. High-molecular
202 weight DNA was extracted from a pool of ~30 pupae from a *C. niger* colony also collected
203 from the Betzet Beach population. A library was constructed and sequenced (CLR
204 sequencing) on a Pacbio Sequel II machine at the Roy J. Carver Biotechnology Center,

205 University of Illinois at Urbana-Champaign. The sequencing yielded 7,031,744 reads, with
206 169 Gbp in total, and an average subread length of 7,227 bp.
207
208 The Pacbio data was assembled into two versions of the genome, Cnig_gn3.1 and
209 Cnig_gn3a.1. Version Cnig_gn3.1 was built more stringently, aiming to include only
210 confidently assembled contigs, while version Cnig_gn3a.1 aimed to map as much of the
211 genome sequence to chromosomes. To avoid technical artifacts in Pacbio sequencing process
212 short (<5 kbp) and long (>50 kbp) reads were excluded from assembly Cnig_gn3.1. Reads
213 were assembled onto contigs using *Canu* version 2.1.1 (Koren, et al. 2017). Contigs of
214 Cnig_gn3a.1 were based on all of PacBio reads. Contigs of both assembly versions were
215 polished by *Pilon* version 1.24 (Walker, et al. 2014) using the short Illumina reads from
216 assembly Cnig_gn1. Polished contigs were anchored to the linkage map (see below) based on
217 markers with a unique *blastn* hits to contigs. Contigs with no map markers were fitted into
218 chromosomes based on the local similarity between the genomes of *C. niger* and *Formica*
219 *selysi* (https://www.ncbi.nlm.nih.gov/assembly/GCA_009859135.1/) if the synteny of the two
220 genomes was colinear in the relevant chromosomal region based on *blast* hits of *C. niger*
221 genetic markers to *F. selysi* chromosomes and *blast* of *F. selysi* genes to *C. niger* contigs.
222 Cnig_gn3.1 assembly chromosomes contain only contigs with unique position and defined
223 orientation. Cnig_gn3a.1 also includes contigs with non-unique mapping and/or unknown
224 orientation. Contigs with several (up to 5) possible positions were included in Cnig_gn3a.1
225 several times. Therefore, the Cnig_gn3a.1 had greater total length (265 vs. 236 Mbp) and
226 N50 contig size (3.8 vs. 1.1 Mbp) relative to Cnig_gn3.1. The Cnig_gn3a.1 assembly
227 includes 83% of the total assembly length in chromosomes, whereas the Cnig_gn3.1
228 assembly includes 74%.
229
230 Gene annotation was conducted using the *GeMoMa* pipeline (Keilwagen, et al. 2016;
231 Keilwagen, et al. 2018) that incorporates three types of evidence: (1) RNA sequencing of
232 samples from all life stages and castes (NCBI accession PRJNA494690; Yahav and Privman
233 2019) aligned to the genome using *HISAT2* (Kim, et al. 2019); (2) homology to genes
234 annotated in other insect genomes (*Camponotus floridanus, Harpegnathos saltator,*
235 *Linepithema humile, Solenopsis invicta, Pogonomyrmex barbatus, Formica exsecta, Apis*
236 *mellifera, Nasonia vitripennis, Anopheles gambiae, Drosophila melanogaster, Tribolium*
237 *castaneum*); and (3) *ab initio* gene prediction using *Augustus* (Stanke, et al. 2006). The
238 annotation resulted in 14,889 and 16,260 putative protein-coding genes for Cnig_gn3.1 and

239 Cnig_gn3a.1, respectively. Functional annotation of genes was assigned based on homology
240 using the *Trinotate* tool (version 3.0.2; https://github.com/Trinotate/Trinotate.github.io/wiki)
241 from the *Trinity* pipeline (Grabherr, et al. 2011), which uses *blast* against the Swissprot
242 protein database and assigns genes to functional categories of the Gene Ontology (GO)
243 database (Harris, et al. 2004). Out of the 16,260 genes in Cnig_gn3a.1, 10,268 genes had
244 Swissprot hits, and GO terms were assigned 10,186 genes. To identify candidate genes in
245 QTLs, we searched both Swissprot gene description and GO terms for keywords related to
246 molecular functions and biological processes implicated in CHC synthesis, including fatty
247 acid metabolism and development and function of the cuticle.
248
249 **RAD sequencing**
250 DNA was extracted individually from each sample using TRIzol® (Invitrogen Life
251 Technologies) and kept at -20°C. Quantity and quality of extracts were determined by
252 NanoDrop™ 2000 Spectrophotometer and run on gel to verify consistency in DNA yield,
253 degradation, and phenol/salts residues. Reduced representation genomic libraries were
254 constructed according to a modified double-digest Restriction-site Associated DNA
255 sequencing (ddRAD-seq) protocol based on Brelsford, et al. (2016). Briefly, DNA was
256 digested by two different restriction enzymes (EcoR1 rare-cutter; Mse1 frequent-cutter) and
257 ligated to barcoded adaptors for multiplexing. Amplified products were pooled and purified
258 (Beckman Coulter, Agencourt® AMPure® XP) so that 200-700bp fragments would be
259 selected. The library was sequenced in one lane of a paired-end, 150bp reads, on a
260 HiSeq4000 Illumina sequencer.
261
262 The number of raw reads totaled in 693,969,414 of which 662,096,312 were associated with a
263 sample (barcoded). The reads were quality checked and those in which sequencing score
264 drops below an average phred score of 10 (i.e. 1 in 10 chance for sequencing error) in a
265 sliding window of 22 bases were removed using *process_radtags* from the *STACKS* pipeline
266 (Catchen, et al. 2011). After this initial filtering a total of 630,112,110 reads remain, which is
267 expected to give an average coverage of ~20X (based on 67,687 positions of the EcoRI
268 restriction sites in the reference genome).
269
270 Reads were aligned to the reference genome using *Bowtie2* (Langmead and Salzberg 2012).
271 Aligned reads that have more than 2 mismatches to the reference genome and reads that had

272 multiple mappings were removed. The number of aligned reads retaind totaled 431,321,204,
273 62% of the original number of reads.
274
275 The aligned reads were analyzed together to create a catalogue of 88,439 SNPs using the
276 *STACKS* pipeline, with an average of 14 reads covering a SNP per sample. These SNPs were
277 filtered using custom perl scripts to remove loci with more than three heterozygous genotypes
278 (not expected for haploid samples) or less than four samples with the rare allele (the brothers
279 are expected to have roughly 50% of each of the maternal alleles) or more than 50% missing
280 data. 7,769 SNPs passed this filtering.
281
282 **Linkage map**
283 The program *MultiPoint ultra-dense* (Ronin, et al. 2017) was used to construct a linkage map,
284 based on the genotypes of 7769 SNPs in the 117 samples. These loci were further filtered by
285 *MultiPoint* for segregation distortion (based on Chi^2 test to reject the null hypothesis of 1:1
286 segregation with $p < 0.005$) and for missing data (no more than 23 missing genotypes),
287 leaving 5466 SNPs. Clustering of linkage groups resulted in 27 clusters of SNP markers
288 using a recombination frequency threshold of 0.28. Skeleton markers were selected in every
289 linkage group (giving priority to "twin markers") and ordered by an "evolutionary strategy"
290 heuristic optimization, as implemented in *MultiPoint*. Ordering was assessed by jackknife
291 resampling. We manually inspected the linkage groups and applied methods implemented in
292 *MultiPoint* to merge linkage groups, exclude markers violating local map monotonicity, and
293 complement the map by adding suitable singleton markers, resulting in a final map consisting
294 of 26 linkage groups. To verify that the map length is not inflated by marker ordering errors,
295 we checked the order by jackknifing and observed a very nearly perfect support across the
296 map (i.e., 100% jackknife repeats gave identical order among each marker neighborhood
297 along the linkage group length).
298
299 **QTL mapping**
300 The program *MultiQTL* (Korol, et al. 1995) was used to map 32 quantitative traits of 81
301 samples, including 31 specific CHC traits and the total CHC amount. We defined the 31
302 specific CHC traits as relative amounts, while total CHC is an absolute amount in ng. Many
303 of the traits had skewed, non-normal distributions, so we removed outliers and applied log-
304 transformation where necessary. We first identified QTLs using the SIM (Single Interval
305 Mapping) model in *MultiQTL*, which calculates a LOD score for every trait on each

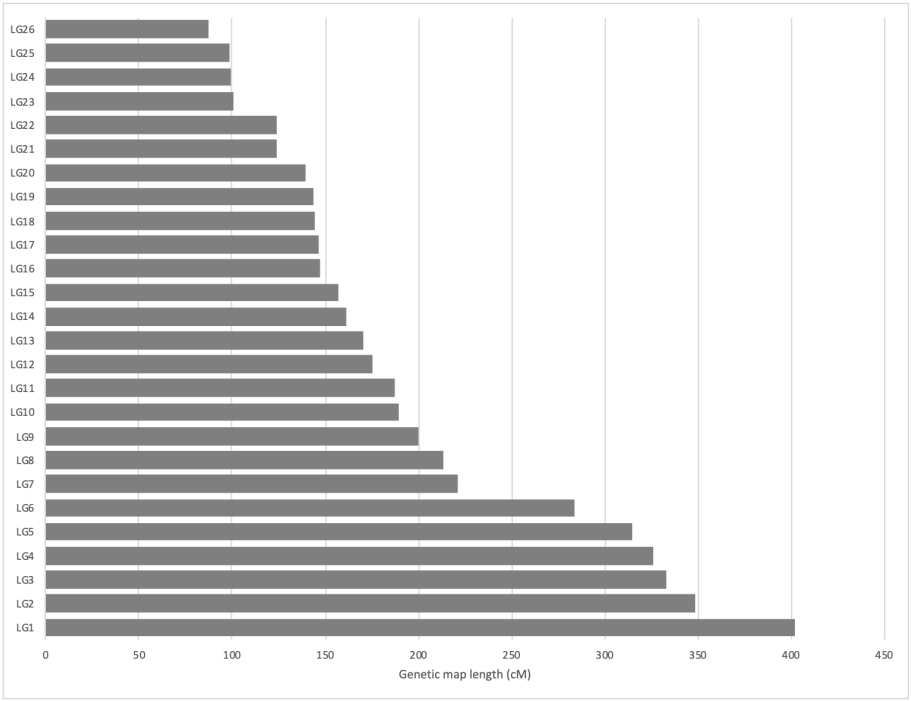
306 chromosome. For every putative QTL with LOD score greater than 1.8, we conducted a
307 permutation test to obtain a *p*-value for the significance of association between genotypes and
308 phenotypes. QTLs with *p*-value < 0.1 were carried on to the next step. For traits with two
309 peaks of the LOD scores on a chromosome, we used the linked-QTL model to test the
310 hypothesis of having two QTLs on the same chromosome (likelihood ratio test). For traits
311 that had candidate QTLs on more than one chromosome, we applied the MIM (multi-interval
312 mapping) model to all chromosomes identified in the SIM analysis, which generally results in
313 higher LOD scores. The permutation test was repeated for every QTL using the MIM model.
314 The SIM, linked-QTL, or MIM models also provided an estimate of the PEV (percent
315 explained variation) for each QTL. We identified the location of each QTL by the peak of the
316 LOD score, and its 95% confidence interval was defined as the region around the maximum,
317 where the LOD score is greater than the maximum score minus 1.5 points. We applied the
318 false discovery rate (FDR) method implemented in R (Benjamini and Hochberg 1995; Team
319 2021) to correct for multiple testing (32 traits by 26 chromosomes = 832 tests). We mapped
320 the genetic position of the confidence interval of each QTL location to the genome assembly
321 relative to the physical positions of genetic map markers. We then listed the annotated genes
322 that reside within the scaffold regions contained in each QTL interval in assembly
323 Cnig_gn3a.1.
324

325 ## Results

326 **Reference genome and linkage map**

327 As a basis for QTL mapping, we generated a new genome assembly for *C. niger* using high-
328 coverage Pacbio sequencing that yielded improved contiguity compared to the first draft
329 (N50=1.08Mb). We then constructed a linkage map using the genotypes of a cohort of 117
330 haploid brothers. We used RAD sequencing to genotype 7769 SNP loci that clustered into 26
331 linkage groups, which is in agreement with the chromosome number in *Cataglyphis*
332 *hispanica*, where one lineage has 26 chromosomes while the other has 27 (Darras, et al.
333 2022). We anchored the assembly contigs to the linkage map, yielding 26 chromosomal
334 sequences that together encompass 83% of the total assembly length. The total map length
335 was 5041cM, which is exceptional for animal species. Some of the chromosomes have a very
336 long genetic map length: chromosome 1 is 402cM long and five additional chromosomes are
337 longer than 280cM (Fig. 1). Given an estimated physical genome size of 236Mb (the total
338 assembly length), we estimated the genome-wide average recombination rate at 21.3 cM/Mb.

339 This rate is similar to the record for animals that is held by honey bees, with estimates
340 ranging between 19-22 cM/Mb (Beye, et al. 2006; Solignac, et al. 2007; Meznar, et al. 2010;
341 Korol and Rybnikov 2024). The total physical and genetic lengths of the *C. niger* genome are
342 both greater than those of *Apis mellifera*: 236Mb and 5041cM vs. 187Mb and 4,114cM
343 (Solignac, et al. 2007).



344

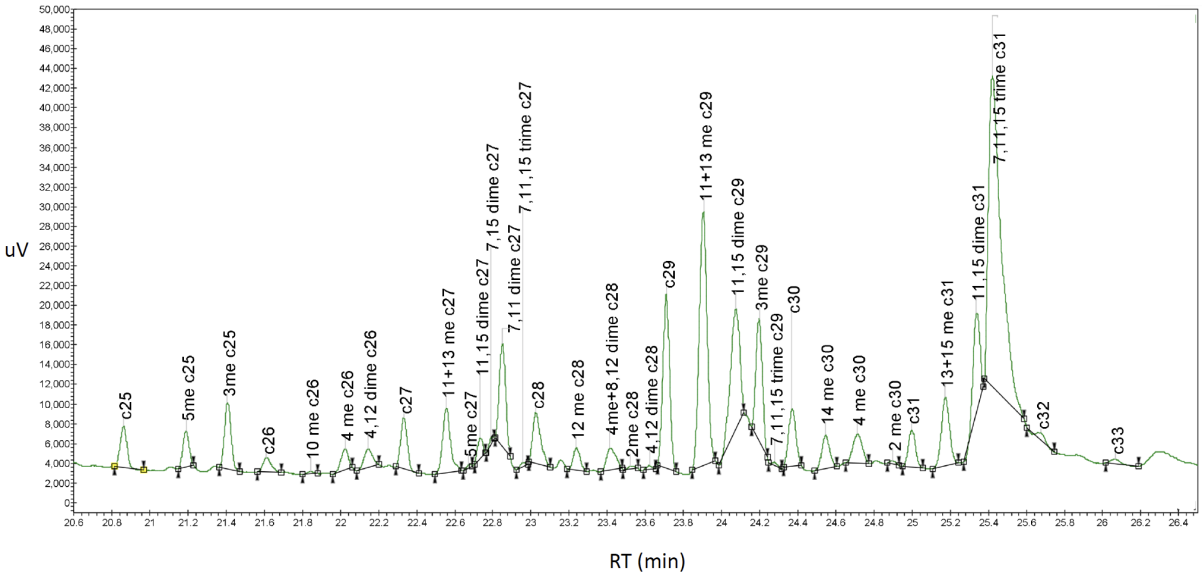
345 **Figure 1:** Distribution of genetic map lengths among the 26 chromosomes of *C. niger*.

346

347 **Chemical analysis of CHCs**

348 We analyzed the cuticular extracts of the same brothers used for linkage map construction,
349 which were treated as quantitative traits for QTL mapping. Chemical analysis by GC/MS was
350 successful for 81 out of the 117 samples, identifying 32 differentiated peaks in the
351 chromatogram. The raw data from the chemical analysis are provided in Supp. File S1. We
352 were able to identify 36 CHCs in 32 peaks (four peaks consisted of two CHCs that failed to
353 separate). These were saturated alkanes with carbon chain length ranging from C25 to C31,
354 many of them having a methyl group, and some with two and even three methyl branches.
355 We did not observe any unsaturated alkenes. Note that this means that our list of candidate
356 genes for CHC synthesis will not include fatty acid desaturases, which are known from the
357 biosynthetic pathway of alkenes in other insects. The quantitative traits for QTL analysis
358 were defined as the proportion of each CHC out of the total CHC mixture. We also included
359 the total CHC amount as an additional trait, which varied from 228 to 4825ng. There was
360 extensive variation both among and within traits. The most abundant peak contained 11Me-

361 and 13Me-C29, and the second most abundant was C29. These two peaks had averages and
362 standard deviations of 226±122ng and 168±121ng, respectively. Conversely, the least
363 abundant CHC, 5Me-C27, was only 1.1±1.8ng, and undetectable in 34 samples. Generally,
364 the distributions of all CHCs were asymmetrical, with many low values and a long right-hand
365 tail (Supp. Fig. S1), which required transformation and removal of outliers before QTL
366 analysis (see Methods). One of these peaks (7,11,15TriMe-C31) was excluded from the
367 analysis because its distribution was extremely non-normal, even after log transformation.
368 Finally, 31 CHC traits were carried forward to the QTL mapping analysis.



369

370 **Figure 2:** A chromatogram (uV absorption vs. retention time; RT) of the non-polar fraction (RT 20.5-
371 26.5) of cuticular extracts from a *C. niger* male. Marked peaks were annotated by mass spectrometry.
372

373 **QTLs associated with variation in CHCs**

374 Each CHC (a peak in the GC chromatograms) was treated as a quantitative trait and tested for
375 association with the map markers (SNPs) in a QTL mapping analysis. The analysis included
376 the relative amounts of 31 CHCs, and the absolute total amount of CHCs as an additional
377 trait. A preliminary analysis identified 62 candidate genomic intervals with LOD scores
378 greater than 1.8. Each candidate was evaluated by permutation analysis, which yielded *p*-
379 values that were then corrected for multiple testing to control the false discovery rate (FDR).
380 This analysis resulted in 20 QTLs (FDR < 10%) for nine traits, including one QTL for total
381 CHC quantity (Table 1). The location and 95% confidence interval of each QTL was
382 estimated based on the LOD function (Figure 3).
383
384

385 **Table 1: QTLs underlying variation in CHCs**

CHC	LOD	Location (cM)	PEV (%)	*p*-value	*q*-value	MIM PEV(%)
C25	6.59*	**Chr6:** 0-43.2 **Chr6:** 76.9-88.8	19.4*	0.001*	0.069*	29.7
C26	4.215	**Chr6:** 56.4-67.4	16.2	0.001	0.069	46.1
	4.168	**Chr17:** 61.0-75.5	16.0	0.001	0.023	
	3.590	**Chr18:** 0-36.5	13.8	0.001	0.062	
C27	3.899	**Chr17:** 87.0-111.9	17.5	0.002	0.078	35.4
	3.981	**Chr18:** 102.1-113.5	18.0	0.001	0.023	
7,11,15TriMe-C27	4.139	**Chr15:** 122.9-154.7	15.0	0.001	0.064	49.4
	4.810	**Chr3:** 207.6-228.1	13.7	0.001	0.023	
	4.120	**Chr6:** 191.7-240.7	12.8	0.001	0.023	
C28	4.761	**Chr10:** 97.9-108.1	14.4	0.001	0.023	57.2%
	6.134	**Chr4:** 8.9-33.0	21.9	0.001	0.023	
	3.910	**Chr5:** 283.0-325.2	13.8	0.002	0.092	
4,12DiMe-C28	4.31*	**Chr20:** 5.0-28.5 **Chr20:** 57.9-84.4	24.9*	0.001*	0.023*	N/A
2Me-C30	3.823	**Chr21:** 10.9-34.2	14.1	0.001	0.064	46.1%
	4.083	**Chr13:** 17.3-31.7	15.6	0.001	0.023	
	3.838	**Chr4:** 0-17.2	16.4	0.002	0.092	
13+15Me-C31	4.837	**Chr18:** 105.9-120.3	25.4	0.001	0.023	N/A
Total CHC	3.886	**Chr15:** 89.3-102.5	15.1	0.001	0.064	44.0%

386 Results from the MIM model, showing only QTLs that passed the 10% FDR cutoff. The reported
387 MIM PEV is based on the analysis of all chromosomes for each trait. The full results for all tested
388 candidate QTLs are reported in Table S1, which includes results that did not pass the significance
389 threshold in the permutation test. LOD = log-of-odds score; Location = confidence interval of QTL
390 location; *q*-value = FDR corrected *p*-value; PEV = percent explained variation; MIM = multi-interval
391 model summarizing all QTLs for a given CHC over different chromosomes.
392 * Two QTLs for the same CHC on the same chromosome evaluated using the linked-QTL model of
393 *MultiQTL*, which gives the LOD, PEV, *p*-value, and *q*-value are for both QTLs together.
394

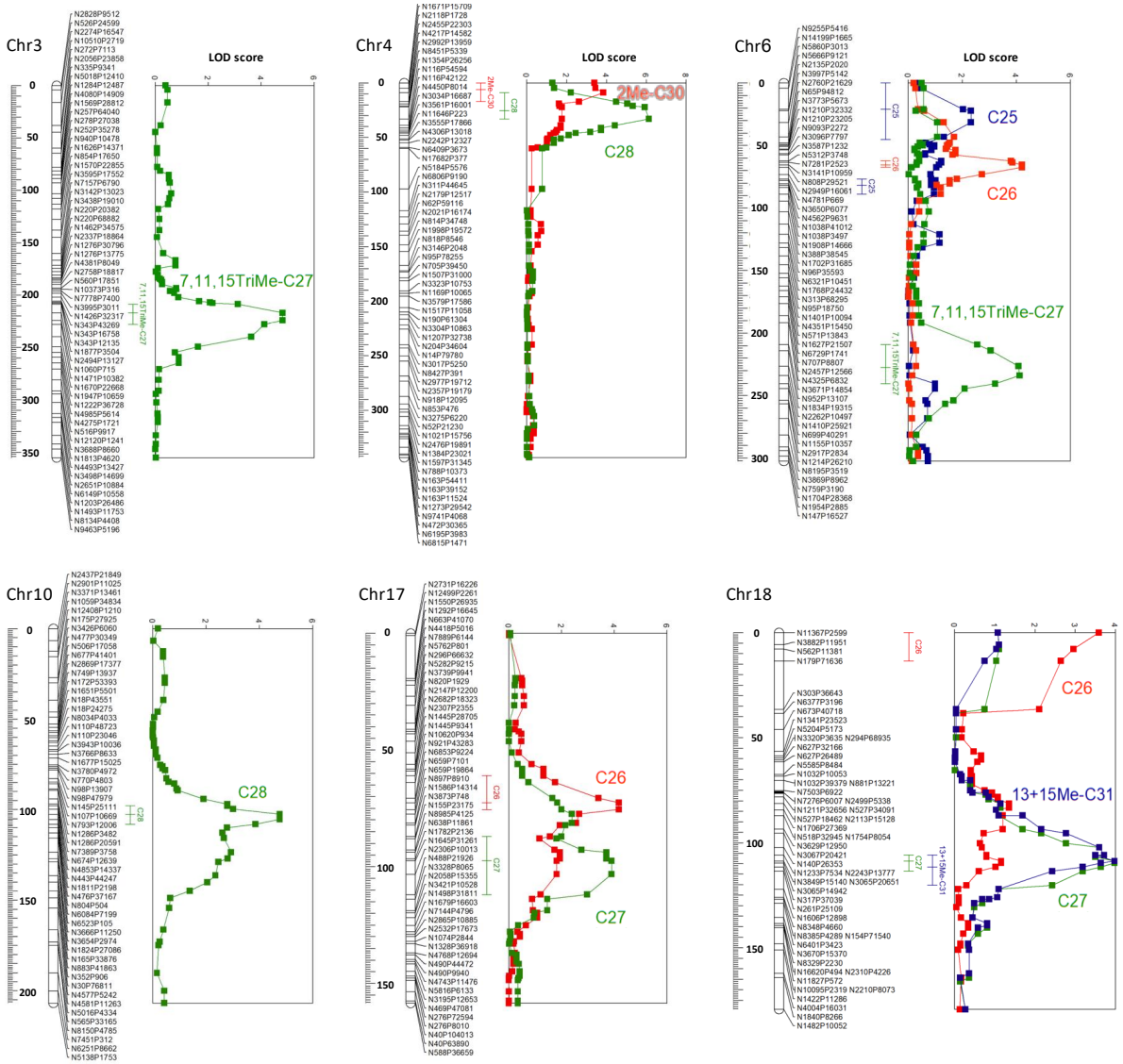


Figure 3: Location of CHC QTLs on six of the *C. niger* chromosomes. Map markers are listed with distances in cM along each linkage group. LOD scores are plotted for each trait, and 95% confidence intervals of QTL locations are indicated.

Interestingly, we did not detect statistically significant QTLs for most CHCs, but when we did find QTLs for a certain CHC we found more than one (Supp. Fig. S2A). Out of the eight CHCs with QTLs, four had three QTLs, five had two QTLs, on only one CHC had a single QTL, which would be unlikely if all CHCs had equal probability to have QTLs (Poisson dispersion test $p = 0.003$). These results indicate that the variation in some CHCs has a greater genetic component than others. QTLs with large effect size are more likely to be discovered with the limited statistical power given our sample size. Thus, other CHCs may still be affected by QTLs with smaller effect sizes, which are below our detection threshold. The portion of the phenotypic variation explained by QTLs (percent explained variation;

409 PEV) for each CHC ranged from 29.7% to 57.2% (Table 1). The PEV for single QTLs varied
410 between 12.8% and 25.4%. The most genetically determined CHC is C28, for which three
411 QTLs had PEV of 14.4%, 13.8%, and 21.9%, and together accounted for 57.2% PEV, which
412 implies that this trait is affected by genetic factors more than by environmental factors.
413
414 The 20 QTLs are located in 11 of the 26 chromosomes of *C. niger*, and the number of QTLs
415 per chromosome was not evenly distributed (Supp. Fig. S2B; Poisson dispersion test p =
416 0.06). In some chromosomes the QTLs for different CHCs appear to be clustered and even
417 overlapping (Fig. 3). For example, QTLs for C28 and 2Me-C30 were found together in the
418 tellomeric region of chromosome 4. Chromosome 18 harbors one locus associated with C26
419 and another associated with C27, 13Me- and 15Me-C31 (the LOD plots are virtually
420 overlapping). Each of the QTLs for C26 and C27 on chromosome 17 are also weakly
421 associated with the other CHC. These results suggest that multiple CHCs may be regulated
422 by the same gene or by different genes in a small genomic region, which could harbor a
423 cluster of co-regulated CHC genes.
424
425 **Candidate genes**
426 Our sample size (n=81) limited the resolution of QTL mapping, resulting in 95% confidence
427 intervals of genetic map length ranging between 10 and 49cM, and physical sizes ranging
428 between 345 and 3,255 Kb. Each interval contained between 21 and 156 putative protein
429 coding genes. Using our data, it is impossible to narrow down these gene lists and identify the
430 loci functionally responsible for the variation in the phenotype. Nevertheless, we noted
431 several candidate genes in these intervals that were implicated in molecular functions and
432 biological processes related to the synthesis of CHCs or to the biology and development of
433 the cuticle (Table 2).
434
435 As discussed in the Introduction, the most obvious candidate genes are ant homologs of
436 enzymes that were extensively studied in model insects and implicated in the biosynthetic
437 pathway of long-chain fatty acids, which are the precursors of CHCs. These include fatty acid
438 synthases, fatty acid elongases, and fatty acid reductases. One such candidate,
439 Cnig_gene3669 in the QTL for C27 on chromosome 17, is a homolog of *Drosophila* fatty
440 acyl-CoA reductase (FAR) genes. The top blast hit in *Drosophila* is the gene *waterproof*
441 (*wat*; see a gene tree of ant and fly FAR genes in Supp. Fig. S3), which is involved in tracheal
442 development. Knocking out *wat* interfered with gas filling of the tracheal tubes during

443 *Drosophila* embryogenesis (Jaspers, et al. 2014) and *wat* homologs in planthoppers affect
444 CHC quantities (Li, et al. 2019; Li, et al. 2020). Generally, FAR genes are responsible for one
445 of the last steps in CHC biosynthetic pathways, where fatty-acyl CoA precursors are
446 converted to hydrocarbons. The FAR gene family was implicated in CHC synthesis in
447 *Drosophila* based on their expression in oenocytes and their role in regulation of oenocyte
448 development (Finet, et al. 2019), association with intraspecific CHC variation, and
449 knockdown experiments that affected CHC length (Dembeck, et al. 2015).
450
451 **Table 2: Candidate genes in QTLs underlying variation in CHCs**

QTL	Gene	Ortholog	Description
2Me-C30 chr13	Cnig_gene5601	FABP9_CAEEL	Fatty acid-binding protein homolog 9
2Me-C30_chr21	Cnig_gene2019	LIPI_HUMAN	Lipase member I
4,12DiMe-C28 chr09	Cnig_gene12502	ECH1_RAT	Delta(3,5)-Delta(2,4)-dienoyl-CoA isomerase, mitochondrial (ECO:0000305)
7,11,15TriMe-C27 chr06	Cnig_gene13278	PEX2_RAT	Peroxisome biogenesis factor 2
7,11,15TriMe-C27 chr15	Cnig_gene4696	ACADV_HUMAN	Very long-chain specific acyl-CoA dehydrogenase, mitochondrial (ECO:0000305|PubMed:7668252)
C26 chr18	Cnig_gene6578	ELO3_CAEEL	Putative fatty acid elongation protein 3
C27 chr17	Cnig_gene3669	WAT_DROME	Fatty acyl-CoA reductase wat (ECO:0000305)
4,12DiMe-C28 chr09	Cnig_gene12805	DHGL_DROME	Glucose dehydrogenase [FAD, quinone]
7,11,15TriMe-C27 chr15	Cnig_gene4695	KAPR1_DROME	cAMP-dependent protein kinase type I regulatory subunit
Total chr15	Cnig_gene4381	FABD_DROME	Probable malonyl-CoA-acyl carrier protein transacylase, mitochondrial

452
453 Another candidate, Cnig_gene6578 in the QTL for C26 on chromosome 18, is a homolog of
454 *Drosophila* long-chain fatty acid elongase genes. This gene family takes part in the CHC
455 synthesis pathway, in elongating the chain length of fatty acids initially synthesized by fatty
456 acid synthases beyond 20 carbons (Blomquist and Bagneres 2010). Knockdown of such
457 elongases in *Drosophila* and in the brown planthopper *Nilaparvata lugens* affected
458 abundances of different CHCs (Holze, et al. 2021). The QTL for 2Me-C30 on chromosome
459 21 contains a homolog of human lipase member I, which hydrolyzes phosphatidic acid (PA)
460 to 2-acyl lysophosphatidic acid (LPA; a potent bioactive lipid mediator) and a fatty acid
461 (Hiramatsu, et al. 2003). While the function of the insect homolog is unknown, it is a

promising candidate in lipid metabolism, which may be involved in the synthesis of CHC precursors.

Additional candidate genes of interest may be implicated in the biology and development of the cuticle. For example, the QTL for 4,12DiMe-C28 on chromosome 9 contains Cnig_gene12805, a homolog of the *Drosophila* glucose dehydrogenase gene *Gld*, which was implicated in cuticle development and suggested to be involved in modification of the puparium cuticle to facilitate eclosion (Cavener, et al. 1986).

Discussion

Here we applied a genomic mapping approach to test for genetic effects on CHCs, which serve as major chemical cues for recognition and other functions in social insects. We analyzed CHC profiles and genomic polymorphism in a sample of 81 haploid brothers from the desert ant *Cataglyphis niger*. We report associations between 19 QTLs and 8 out of 31 CHCs tested, and one locus that is associated with the total amount of CHCs. We identified rather large effect sizes that each explain between 13% and 25% of the variation in a given CHC. Between 30% and 57% of the variation of the eight affected CHCs was explained by the combined genetic effects. Our study opens a window to the genomic architecture underlying CHC variation in a social insect species.

Our results adds to a large volume of literature on determinants of CHC profiles. Previous studies demonstrated large environmental effects by various factors, including diet (Liang and Silverman 2000), pathogens (Tranter, et al. 2015; Leclerc and Detrain 2016), and abiotic environmental factors such as temperature and humidity (Silva, et al. 2024). Incorporating our results allows a comprehensive understanding of the mechanisms shaping CHC profiles by a combination of genetic and environmental effects. Our results reveal large variation in the genetic effects on different CHCs, with up to 57% of the variation of some CHCs explained by the QTLs we identified. CHCs with QTLs are expected to be heritable as suggested by previous studies in multiple ant species (Nowbahari, et al. 1990; Lahav, et al. 1999; Akino, et al. 2004; Ozaki, et al. 2005; van Zweden, et al. 2009; Walsh, et al. 2020). Our results are in line with estimates of narrow-sense heritability (h^2) of individual CHCs in the ant *Monomorium pharaonis* that ranged between 0 and 0.36 (Walsh, et al. 2020). We could not detect any QTLs associated with 23 out of the 31 CHCs in our data. This may be

495 because some CHCs are only affected by small effect size QTLs that could not be detected
496 with our sample size. Generally, our estimates of the total genetic component for any of the
497 studied CHCs should be considered as a lower bound, because they may be affected by
498 additional small effect size QTLs that we could not detect. Nevertheless, our results indicate
499 substantial variation among CHCs in the magnitude of their genetic basis and that for most
500 CHCs the genetic component is smaller than the environmental component. Environmental
501 effects may dominate the cues for nestmate recognition because nestmates share the same
502 abiotic environment of the nest, the same diet, and other biotic effects such as pathogens.
503
504 The eight CHCs with associated QTLs include four straight-chain and four methyl-branched
505 alkanes. This may be surprising considering that methyl-branched CHCs have greater
506 information content, and thus are expected to be under greater selection as recognition cues,
507 and more likely to evolve a genetic basis. This hypothesis was supported by some studies that
508 reported stronger differentiation between colonies by methyl-branched vs. straight alkanes in
509 some species (e.g., in *Cataglyphis iberica*; Dahbi, et al. 1996) and stronger direct and indirect
510 genetic effects (e.g., in *Formica rufibarbis*; Van Zweden, et al. 2010). There may be other
511 reasons for the evolution of a genetic basis for the less informative straight-chain alkanes,
512 other than for nestmate recognition purposes. For example, these alkanes may be used for
513 simple, conserved signals that do not require high information content, such as differentiating
514 between foragers and nurses (Greene and Gordon 2003; Martin and Drijfhout 2009b) or to
515 identify a mated queen (Hora, et al. 2008). If alkanes are used for such a conserved function
516 they may evolve a genetic basis to better regulate their expression.
517
518 In this study we used a cohort of male brothers as the mapping population. The motivation
519 for this choice was that males are haploid, which is advantageous for genomic mapping, and
520 are raised in the nest in uniform environmental conditions. Male CHCs are likely to have
521 functional roles for recognition and discrimination in mate choice during mating flights.
522 CHCs were extensively investigated in such roles in other, non-social insects. For example,
523 male offspring from a cross between the wasps *Nasonia vitripennis* and *N. giraulti* were used
524 for a QTL mapping study of CHCs (Niehuis, et al. 2011). CHCs were demonstrated to
525 function as cues for mate choice in this system, which allows females to discriminate
526 between males from different species (Buellesbach, et al. 2013). Nestmate recognition in the
527 worker caste serves its most consequential role for identifying intruders from other colonies.
528 The genetic effects we uncovered in males are likely to also affect CHC profiles of queens

529 and workers, because most CHCs are observed in more than one caste of the same species
530 (Hefetz 2007). Genes involved in the synthesis of these CHCs likely evolved in response to
531 selection pressures on a range of functional roles in all castes. Natural selection on the ability
532 to discriminate against foreign workers may increase genetic polymorphism that contributes
533 to CHC variation, and this polymorphism may also affect variation of the same CHC in males
534 even if this CHC is not under selection in males. Therefore, some of the QTLs we detected in
535 males may actually be functionally important in one of the other castes. Further studies in
536 other castes will provide a more comprehensive understanding of this genetic system and its
537 functional significance.
538
539

540 Acknowledgements

541 We thank Abraham Hefetz for advice and access to equipment for chemical analysis of
542 CHCs. We thank Jessica Purcell and Alan Brelsford for advice on RAD sequencing. This
543 study was funded by Israel Science Foundation grant number 646/15.
544
545

546 Data availability

547 All sequence data was deposited to NCBI: the new genome assembly Cnig_gn3.1 is under
548 NCBI BioProject accession PRJNA1168153 and the raw Pacbio sequencing reads are under
549 SRA accession SRR10302333. The RAD-seq data is under Bioproject accession
550 PRJNA1000554. Chemical data is provided as supplementary materials to this paper.
551
552

553 References

554 Akino T, Yamamura K, Wakamura S, Yamaoka R. 2004. Direct behavioral evidence for
555 hydrocarbons as nestmate recognition cues in Formica japonica (Hymenoptera : Formicidae).
556 Applied Entomology and Zoology 39:381-387.
557 Ayala FJ, Campbell CA. 1974. Frequency-dependent selection. Annual Review of Ecology and
558 Systematics 5:115-138.
559 Benjamini Y, Hochberg Y. 1995. Controlling the false discovery rate: a practical and powerful
560 approach to multiple testing. J R Stat Soc B Met 57:289-300.
561 Beye M, Gattermeier I, Hasselmann M, Gempe T, Schioett M, Baines JF, Schlipalius D, Mougel F,
562 Emore C, Rueppell O, et al. 2006. Exceptionally high levels of recombination across the honey
563 bee genome. Genome Res 16:1339-1344.
564 Blomquist GJ, Bagneres A-G. 2010. Insect hydrocarbons: biology, biochemistry, and chemical
565 ecology: Cambridge University Press.

566 Blomquist GJ, Tittiger C, Jurenka R. 2020. Cuticular hydrocarbons and pheromones of arthropods. In:
567 Wilkes H, editor. Hydrocarbons, oils and lipids: diversity, origin, chemistry and fate. Cham:
568 Springer. p. 213-244.
569 Bonavita-Cougourdan A, Clément J, Lange C. 1987. Nestmate recognition: the role of cuticular
570 hydrocarbons in the ant Camponotus vagus Scop.
571 Brelsford A, Dufresnes C, Perrin N. 2016. High-density sex-specific linkage maps of a European tree
572 frog (Hyla arborea) identify the sex chromosome without information on offspring sex.
573 Heredity (Edinb) 116:177-181.
574 Buellesbach J, Gadau J, Beukeboom LW, Echinger F, Raychoudhury R, Werren JH, Schmitt T. 2013.
575 Cuticular hydrocarbon divergence in the jewel wasp Nasonia: evolutionary shifts in chemical
576 communication channels? Journal of evolutionary biology 26:2467-2478.
577 Cappa F, Petrocelli I, Dani FR, Dapporto L, Giovannini M, Silva-Castellari J, Turillazzi S, Cervo R.
578 2019. Natural biocide disrupts nestmate recognition in honeybees. Scientific reports 9:3171.
579 Catchen JM, Amores A, Hohenlohe P, Cresko W, Postlethwait JH. 2011. Stacks: building and
580 genotyping Loci de novo from short-read sequences. G3 1:171-182.
581 Cavener D, Corbett G, Cox D, Whetten R. 1986. Isolation of the eclosion gene cluster and the
582 developmental expression of the Gld gene in Drosophila melanogaster. The EMBO journal
583 5:2939-2948.
584 Cho S, Huang ZY, Green DR, Smith DR, Zhang J. 2006. Evolution of the complementary sex-
585 determination gene of honey bees: Balancing selection and trans-species polymorphisms.
586 Genome Research 16:1366-1375.
587 d'Ettorre P, Lenoir A. 2010. Nestmate recognition. Ant ecology:194-209.
588 Dahbi A, Cerdá X, Hefetz A, Lenoir A. 1996. Social closure, aggressive behavior, and cuticular
589 hydrocarbon profiles in the polydomous ant Cataglyphis iberica (Hymenoptera, Formicidae).
590 Journal of Chemical Ecology 22:2173-2186.
591 Darras H, de Souza Araujo N, Baudry L, Guiglielmoni N, Lorite P, Marbouty M, Rodriguez F,
592 Arkhipov I, Koszul R, Flot J-F. 2022. Chromosome-level genome assembly and annotation of
593 two lineages of the ant Cataglyphis hispanica: steppingstones towards genomic studies of
594 hybridogenesis and thermal adaptation in desert ants. bioRxiv.
595 Dembeck LM, Böröczky K, Huang W, Schal C, Anholt RR, Mackay TF. 2015. Genetic architecture
596 of natural variation in cuticular hydrocarbon composition in Drosophila melanogaster. Elife
597 4:e09861.
598 Derstine N, Gries R, Zhai H, Jimenez S, Gries G. 2018. Cuticular hydrocarbons determine sex, caste,
599 and nest membership in each of four species of yellowjackets (Hymenoptera: Vespidae).
600 Insectes Sociaux 65:581-591.
601 Dronnet S, Lohou C, Christides J-P, Bagnères A-G. 2006. Cuticular hydrocarbon composition reflects
602 genetic relationship among colonies of the introduced termite Reticulitermes santonensis
603 Feytaud. Journal of Chemical Ecology 32:1027-1042.
604 Ferveur J-F. 2005. Cuticular hydrocarbons: their evolution and roles in Drosophila pheromonal
605 communication. Behavior genetics 35:279-295.
606 Finet C, Slavik K, Pu J, Carroll SB, Chung H. 2019. Birth-and-death evolution of the fatty acyl-CoA
607 reductase (FAR) gene family and diversification of cuticular hydrocarbon synthesis in
608 Drosophila. Genome biology and evolution 11:1541-1551.
609 Grabherr MG, Haas BJ, Yassour M, Levin JZ, Thompson DA, Amit I, Adiconis X, Fan L,
610 Raychowdhury R, Zeng Q. 2011. Trinity: reconstructing a full-length transcriptome without a
611 genome from RNA-Seq data. Nature biotechnology 29:644.
612 Greene MJ, Gordon DM. 2003. Cuticular hydrocarbons inform task decisions. Nature 423:32-32.
613 Harris MA, Clark J, Ireland A, Lomax J, Ashburner M, Foulger R, Eilbeck K, Lewis S, Marshall B,
614 Mungall C, et al. 2004. The Gene Ontology (GO) database and informatics resource. Nucleic
615 acids research 32:D258-261.
616 Hefetz A. 2007. The evolution of hydrocarbon pheromone parsimony in ants (Hymenoptera:
617 Formicidae)—interplay of colony odor uniformity and odor idiosyncrasy. Myrmecological
618 News 10:59-68.
619 Hiramatsu T, Sonoda H, Takanezawa Y, Morikawa R, Ishida M, Kasahara K, Sanai Y, Taguchi R,
620 Aoki J, Arai H. 2003. Biochemical and molecular characterization of two phosphatidic acid-

621 selective phospholipase A1s, mPA-PLA1α and mPA-PLA1β. Journal of Biological Chemistry
622 278:49438-49447.
623 Holldobler B, Michener CD. 1980. Mechanisms of identification and discrimination in social
624 Hymenoptera. In: Markl H, editor. Evolution of social behaviour: hypotheses and empirical
625 tests. p. 35-58.
626 Holze H, Schrader L, Buellesbach J. 2021. Advances in deciphering the genetic basis of insect
627 cuticular hydrocarbon biosynthesis and variation. Heredity 126:219-234.
628 Hora RR, Ionescu-Hirsh A, Simon T, Delabie J, Robert J, Fresneau D, Hefetz A. 2008. Postmating
629 changes in cuticular chemistry and visual appearance in Ectatomma tuberculatum queens
630 (Formicidae: Ectatomminae). Naturwissenschaften 95:55-60.
631 Howard RW, McDaniel C, Nelson DR, Blomquist GJ, Gelbaum LT, Zalkow LH. 1982. Cuticular
632 hydrocarbons of Reticulitermes virginicus (Banks) and their role as potential species-and caste-
633 recognition cues. Journal of Chemical Ecology 8:1227-1239.
634 Hughes AL, Nei M. 1988. Pattern of nucleotide substitution at major histocompatibility complex class
635 I loci reveals overdominant selection. Nature 335:167-170.
636 Jaspers MH, Pflanz R, Riedel D, Kawelke S, Feussner I, Schuh R. 2014. The fatty acyl-CoA reductase
637 Waterproof mediates airway clearance in Drosophila. Developmental biology 385:23-31.
638 Jongepier E, Foitzik S. 2016. Ant recognition cue diversity is higher in the presence of slavemaker
639 ants. Behavioral Ecology 27:304-311.
640 Keilwagen J, Hartung F, Paulini M, Twardziok SO, Grau J. 2018. Combining RNA-seq data and
641 homology-based gene prediction for plants, animals and fungi. BMC bioinformatics 19:1-12.
642 Keilwagen J, Wenk M, Erickson JL, Schattat MH, Grau J, Hartung F. 2016. Using intron position
643 conservation for homology-based gene prediction. Nucleic acids research 44:e89-e89.
644 Kim D, Paggi JM, Park C, Bennett C, Salzberg SL. 2019. Graph-based genome alignment and
645 genotyping with HISAT2 and HISAT-genotype. Nature biotechnology 37:907-915.
646 Kleeberg I, Menzel F, Foitzik S. 2017. The influence of slavemaking lifestyle, caste and sex on
647 chemical profiles in Temnothorax ants: insights into the evolution of cuticular hydrocarbons.
648 Proceedings of the Royal Society B: Biological Sciences 284:20162249.
649 Korol A, Rybnikov S. 2024. Recombination. In. Encyclopedia of Biodiversity: Elsevier. p. 375-396.
650 Korol AB, Ronin YI, Kirzhner VM. 1995. Interval mapping of quantitative trait loci employing
651 correlated trait complexes. Genetics 140:1137-1147.
652 Lahav S, Soroker V, Hefetz A, Vander Meer RK. 1999. Direct behavioral evidence for hydrocarbons
653 as ant recognition discriminators. Naturwissenschaften 86:246-249.
654 Lajmi A, Cohen P, Lee C-C, Frenkel Z, Pellen Y, Privman E. 2024. Repeated evolution of supergenes
655 on an ancient social chromosome. bioRxiv:2024.2012. 2001.626239.
656 Langmead B, Salzberg SL. 2012. Fast gapped-read alignment with Bowtie 2. Nature methods 9:357-
657 359.
658 Leclerc J-B, Detrain C. 2016. Ants detect but do not discriminate diseased workers within their nest.
659 The Science of Nature 103:1-12.
660 Li D-T, Chen X, Wang X-Q, Zhang C-X. 2019. FAR gene enables the brown planthopper to walk and
661 jump on water in paddy field. Science China Life Sciences 62:1521-1531.
662 Li DT, Dai YT, Chen X, Wang XQ, Li ZD, Moussian B, Zhang CX. 2020. Ten fatty acyl-CoA
663 reductase family genes were essential for the survival of the destructive rice pest, Nilaparvata
664 lugens. Pest management science 76:2304-2315.
665 Liang D, Silverman J. 2000. "You are what you eat": diet modifies cuticular hydrocarbons and
666 nestmate recognition in the Argentine ant, Linepithema humile. Naturwissenschaften 87:412-
667 416.
668 Martin S, Drijfhout F. 2009a. A review of ant cuticular hydrocarbons. Journal of Chemical Ecology
669 35:1151-1161.
670 Martin S, Trontti K, Shemilt S, Drijfhout F, Butlin R, Jackson D. 2012. Weak patriline effects are
671 present in the cuticular hydrocarbon profiles of isolated Formica exsecta ants but they disappear
672 in the colony environment. Ecology and evolution 2:2333-2346.
673 Martin SJ, Drijfhout FP. 2009b. Nestmate and task cues are influenced and encoded differently within
674 ant cuticular hydrocarbon profiles. Journal of Chemical Ecology 35:368-374.

Martin SJ, Vitikainen E, Helantera H, Drijfhout FP. 2008. Chemical basis of nest-mate discrimination in the ant Formica exsecta. Proceedings of the Royal Society B-Biological Sciences 275:1271-1278.

Meznar ER, Gadau J, Koeniger N, Rueppell O. 2010. Comparative linkage mapping suggests a high recombination rate in all honeybees. Journal of Heredity 101:S118-S126.

Niehuis O, Bullesbach J, Judson AK, Schmitt T, Gadau J. 2011. Genetics of cuticular hydrocarbon differences between males of the parasitoid wasps Nasonia giraulti and Nasonia vitripennis. Heredity (Edinb) 107:61-70.

Nowbahari E, Lenoir A, Clément J, Lange C, Bagneres A, Joulie C. 1990. Individual, geographical and experimental variation of cuticular hydrocarbons of the ant Cataglyphis cursor (Hymenoptera: Formicidae): their use in nest and subspecies recognition. Biochemical Systematics and Ecology 18:63-73.

Nunes T, Turatti I, Mateus S, Nascimento F, Lopes N, Zucchi R. 2009. Cuticular hydrocarbons in the stingless bee Schwarziana quadripunctata (Hymenoptera, Apidae, Meliponini): differences between colonies, castes and age. Genetics and Molecular Research 8:589-595.

Ozaki M, Hefetz A. 2014. Neural Mechanisms and Information Processing in Recognition Systems. Insects 5:722-741.

Ozaki M, Wada-Katsumata A, Fujikawa K, Iwasaki M, Yokohari F, Satoji Y, Nisimura T, Yamaoka R. 2005. Ant nestmate and non-nestmate discrimination by a chemosensory sensillum. Science 309:311-314.

Ronin YI, Mester DI, Minkov DG, Akhunov E, Korol AB. 2017. Building ultra-high-density linkage maps based on efficient filtering of trustable markers. Genetics 206:1285-1295.

Silva KL, Batista NR, Antonialli-Junior WF. 2024. Effect of Temperature on Nestmate Recognition in the Ant Odontomachus chelifer. Sociobiology 71:e9944-e9944.

Singer TL. 1998. Roles of hydrocarbons in the recognition systems of insects. American Zoologist 38:394-405.

Solignac M, Mougel F, Vautrin D, Monnerot M, Cornuet J-M. 2007. A third-generation microsatellite-based linkage map of the honey bee, Apis mellifera, and its comparison with the sequence-based physical map. Genome Biol 8:1-14.

Sprenger PP, Menzel F. 2020. Cuticular hydrocarbons in ants (Hymenoptera: Formicidae) and other insects: how and why they differ among individuals, colonies, and species. Myrmecological News 30.

Stanke M, Keller O, Gunduz I, Hayes A, Waack S, Morgenstern B. 2006. AUGUSTUS: ab initio prediction of alternative transcripts. Nucleic acids research 34:W435-W439.

Stuart RJ, Herbers JM. 2000. Nest mate recognition in ants with complex colonies: within-and between-population variation. Behavioral Ecology 11:676-685.

Sturgis SJ, Gordon DM. 2012. Nestmate recognition in ants (Hymenoptera: Formicidae): a review. Myrmecological News 16:101-110.

Team RC. 2021. R: A language and environment for statistical computing. Vienna, Austria: R Foundation for Statistical Computing.

Thomas M, Simmons L. 2008. Cuticular hydrocarbons are heritable in the cricket Teleogryllus oceanicus. Journal of evolutionary biology 21:801-806.

Tranter C, LeFevre L, Evison SE, Hughes WO. 2015. Threat detection: contextual recognition and response to parasites by ants. Behavioral Ecology 26:396-405.

Van Zweden J, Brask JB, Christensen JH, Boomsma JJ, Linksvayer TA, D'ettorre P. 2010. Blending of heritable recognition cues among ant nestmates creates distinct colony gestalt odours but prevents within-colony nepotism. Journal of evolutionary biology 23:1498-1508.

van Zweden JS, Dreier S, d'Ettorre P. 2009. Disentangling environmental and heritable nestmate recognition cues in a carpenter ant. Journal of insect physiology 55:159-164.

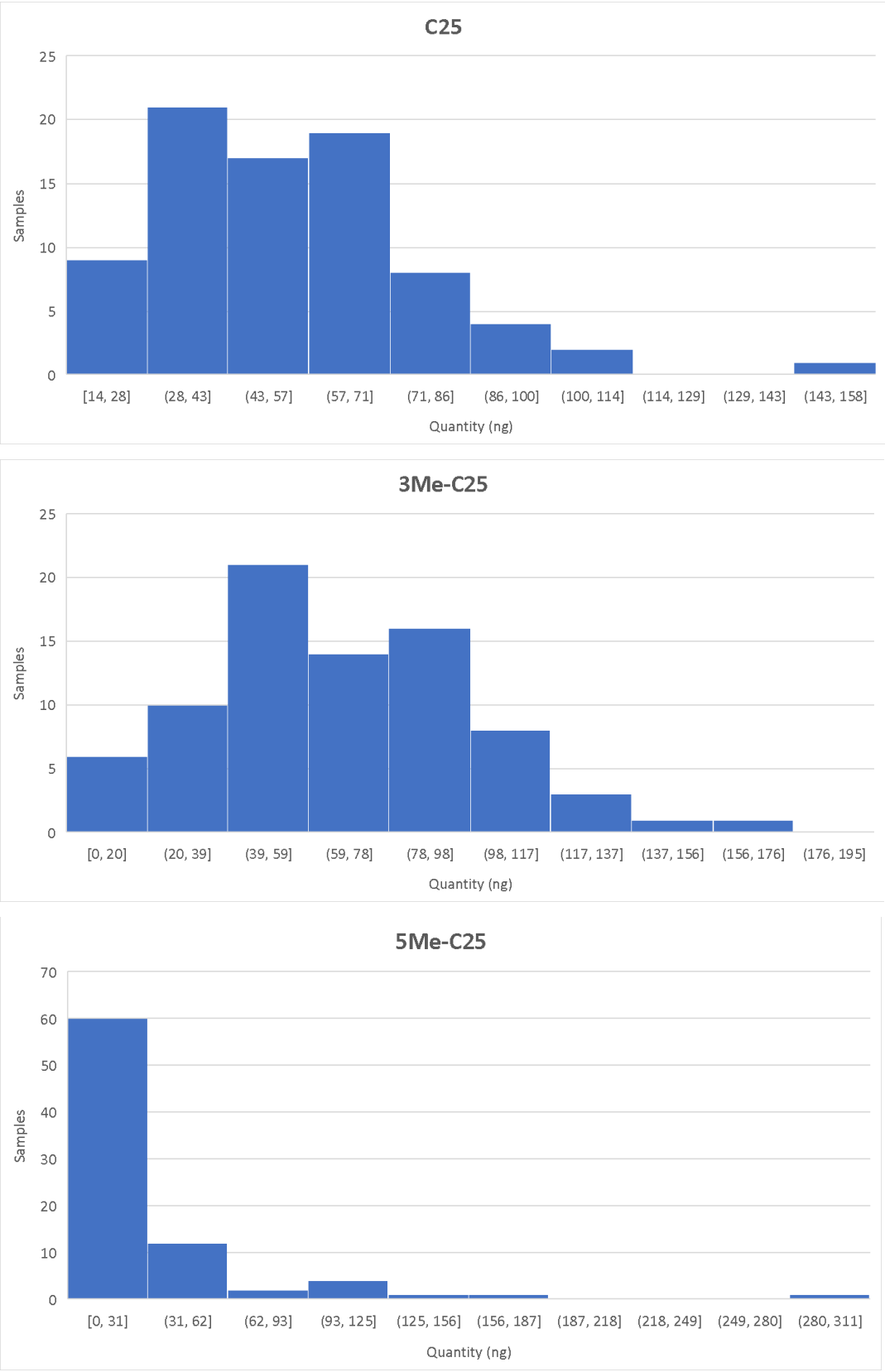
Walsh J, Pontieri L, d'Ettorre P, Linksvayer TA. 2020. Ant cuticular hydrocarbons are heritable and associated with variation in colony productivity. Proc Biol Sci 287:20201029.

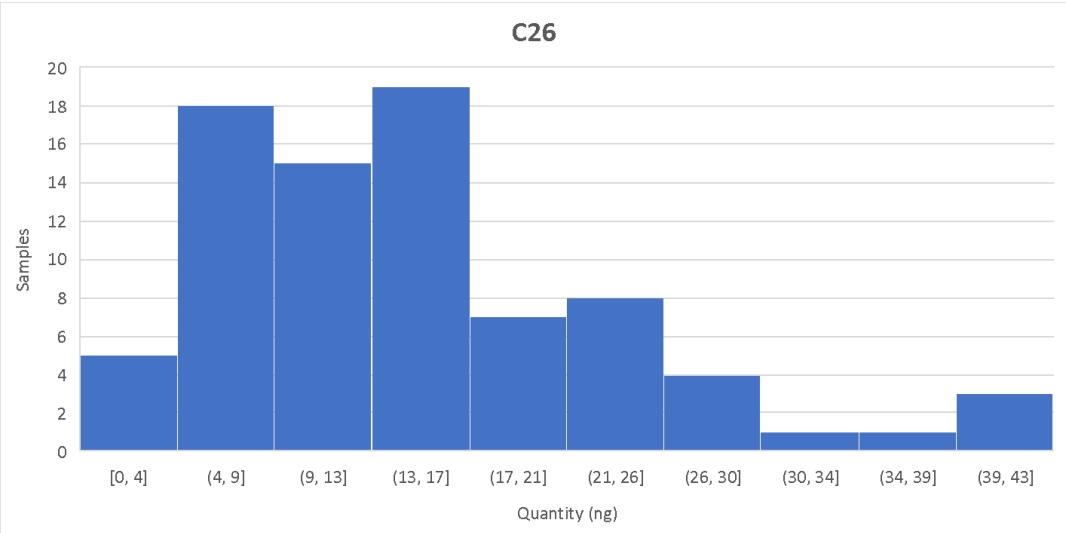
Yahav T, Privman E. 2019. A comparative analysis of methods for de novo assembly of hymenopteran genomes using either haploid or diploid samples. Sci Rep 9:6480.

729

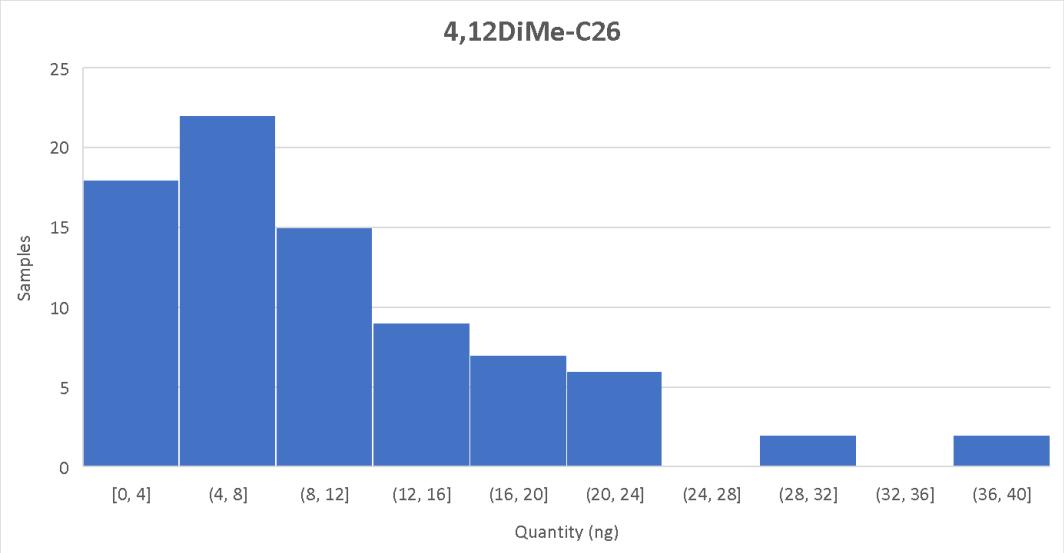
Supplementary

Supplementary Figure S1: CHC distributions

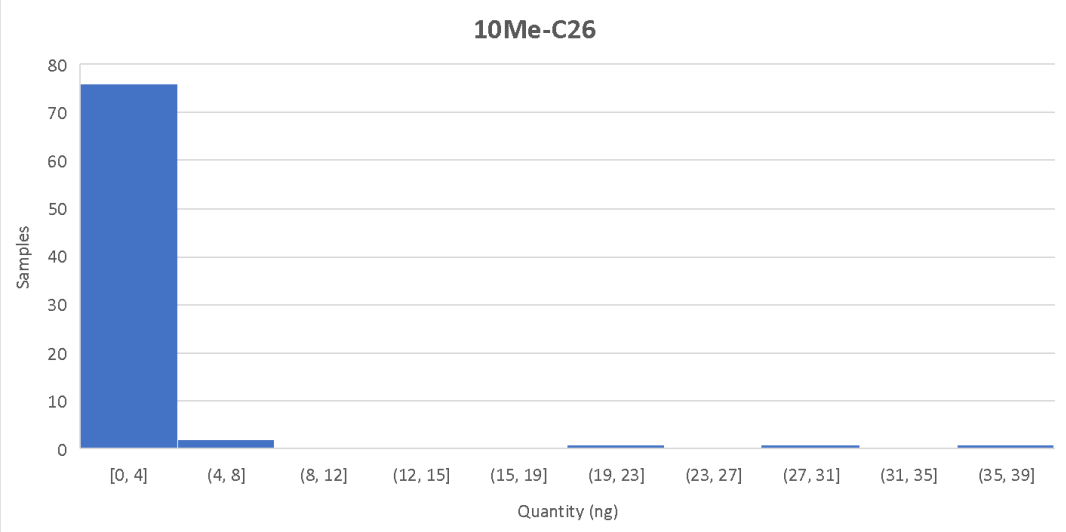



C26
Samples
20
18
16
14
12
10
8
6
4
2
0
[0, 4]
(4, 9]
(9, 13]
(13, 17]
(17, 21]
(21, 26]
(26, 30]
(30, 34]
(34, 39]
(39, 43]
Quantity (ng)

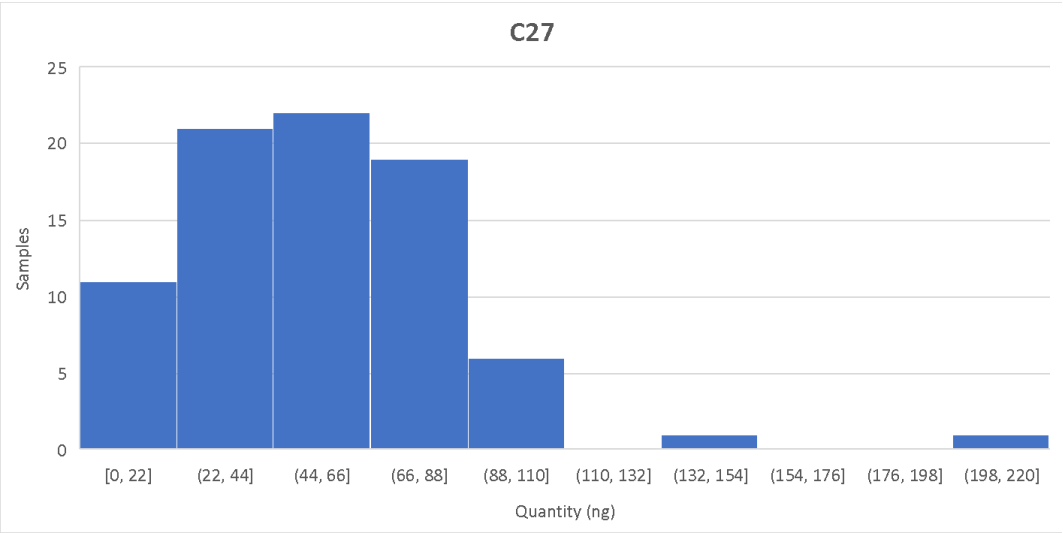
735

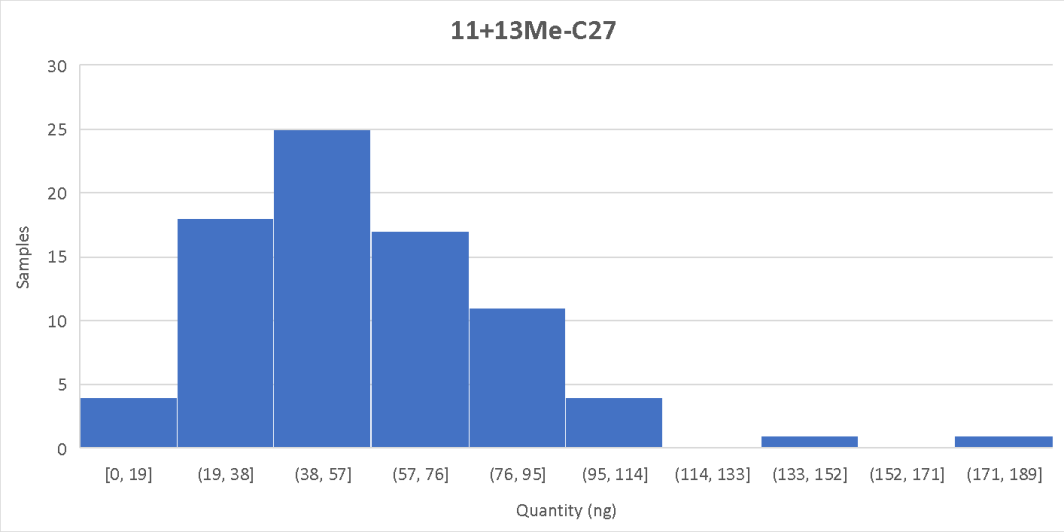

4,12DiMe-C26
Samples
25
20
15
10
5
0
[0, 4]
(4, 8]
(8, 12]
(12, 16]
(16, 20]
(20, 24]
(24, 28]
(28, 32]
(32, 36]
(36, 40]
Quantity (ng)

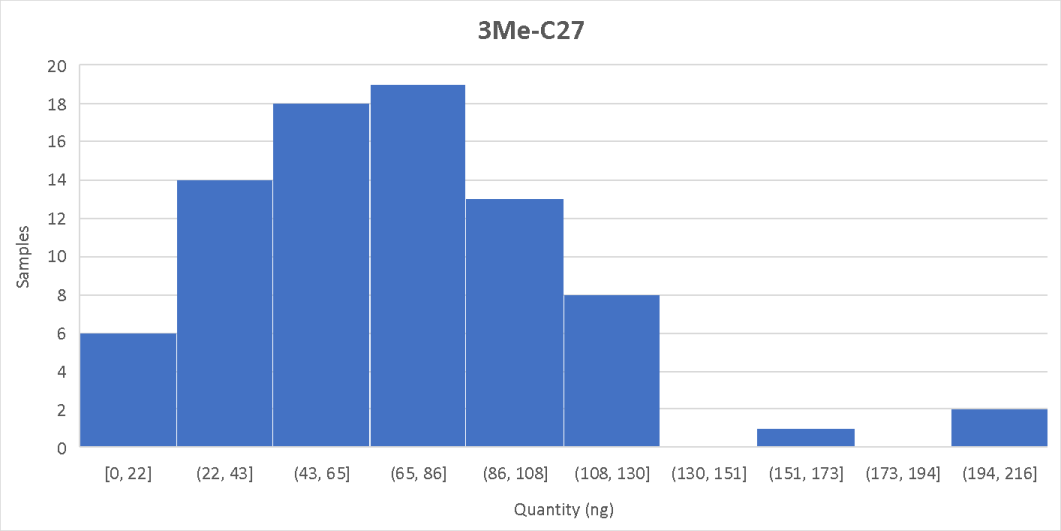
736

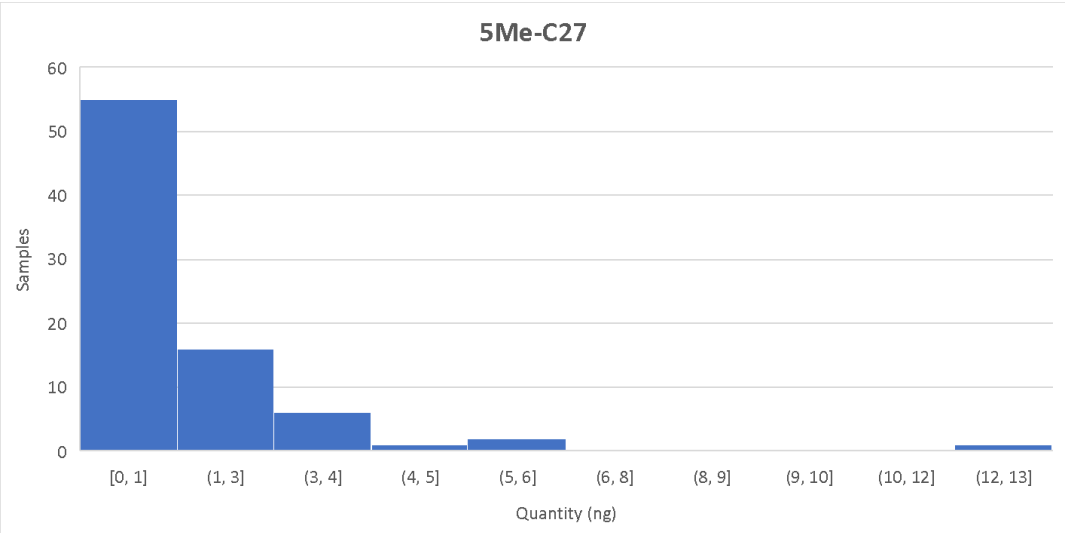

10Me-C26
Samples
80
70
60
50
40
30
20
10
0
[0, 4]
(4, 8]
(8, 12]
(12, 15]
(15, 19]
(19, 23]
(23, 27]
(27, 31]
(31, 35]
(35, 39]
Quantity (ng)

737

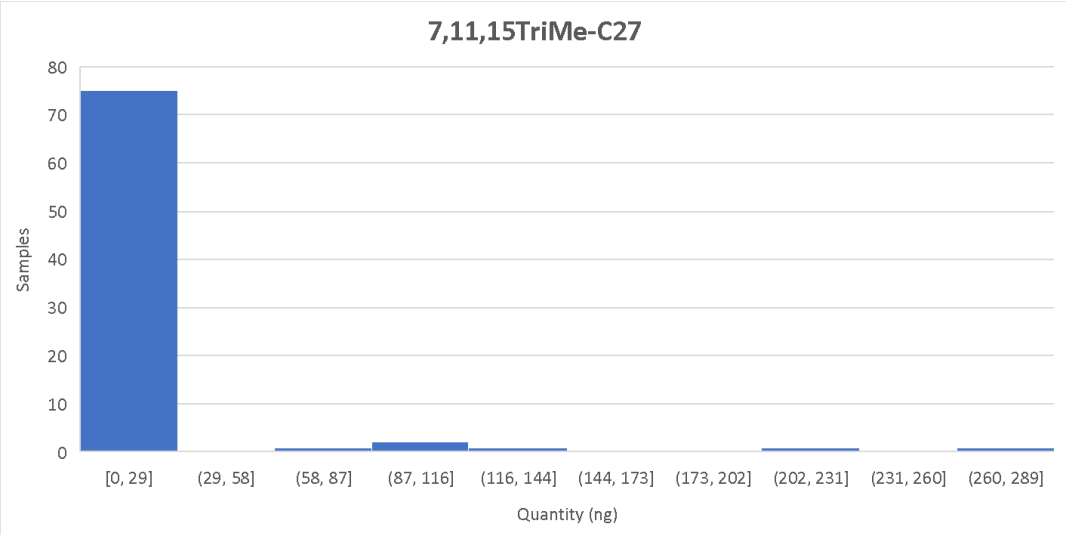

C27
Samples
25
20
15
10
5
0
[0, 22]
(22, 44]
(44, 66]
(66, 88]
(88, 110]
(110, 132]
(132, 154]
(154, 176]
(176, 198]
(198, 220]
Quantity (ng)


11+13Me-C27
Samples
30
25
20
15
10
5
0
[0, 19]
(19, 38]
(38, 57]
(57, 76]
(76, 95]
(95, 114]
(114, 133]
(133, 152]
(152, 171]
(171, 189]
Quantity (ng)

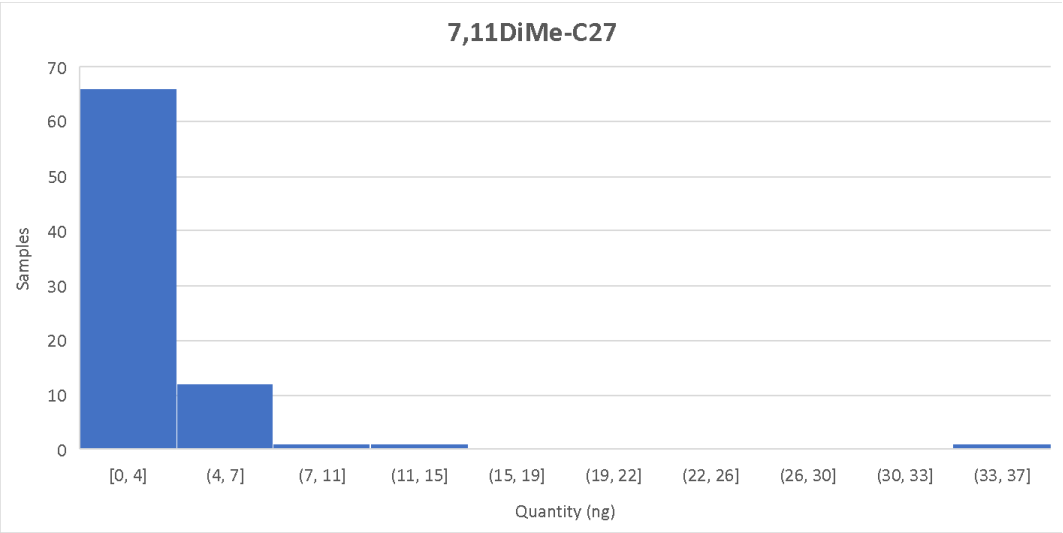

3Me-C27
Samples
20
18
16
14
12
10
8
6
4
2
0
[0, 22]
(22, 43]
(43, 65]
(65, 86]
(86, 108]
(108, 130]
(130, 151]
(151, 173]
(173, 194]
(194, 216]
Quantity (ng)


5Me-C27
Samples
60
50
40
30
20
10
0
[0, 1]
(1, 3]
(3, 4]
(4, 5]
(5, 6]
(6, 8]
(8, 9]
(9, 10]
(10, 12]
(12, 13]
Quantity (ng)

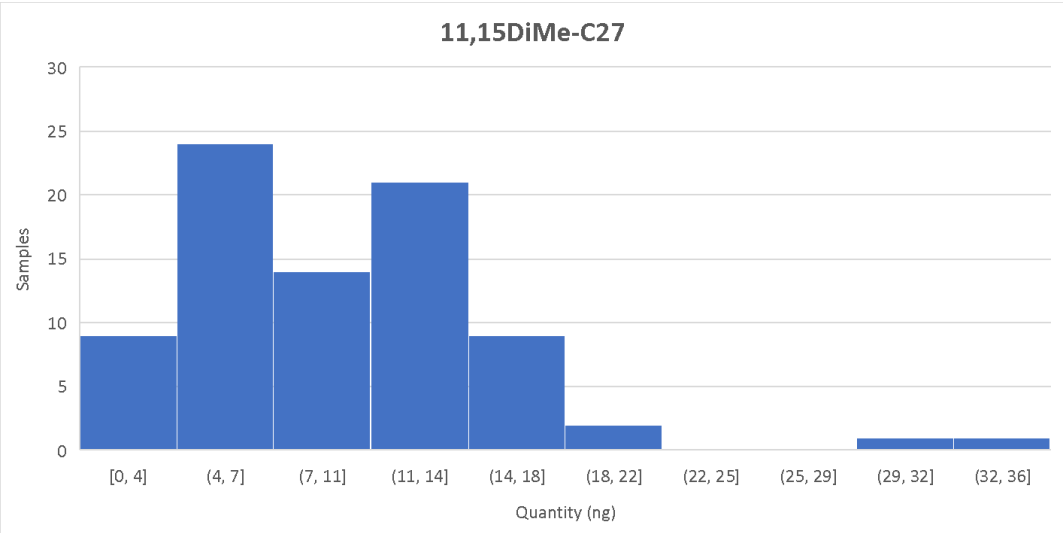
741


7,11,15TriMe-C27
Samples
80
70
60
50
40
30
20
10
0
[0, 29]
(29, 58]
(58, 87]
(87, 116]
(116, 144]
(144, 173]
(173, 202]
(202, 231]
(231, 260]
(260, 289]
Quantity (ng)

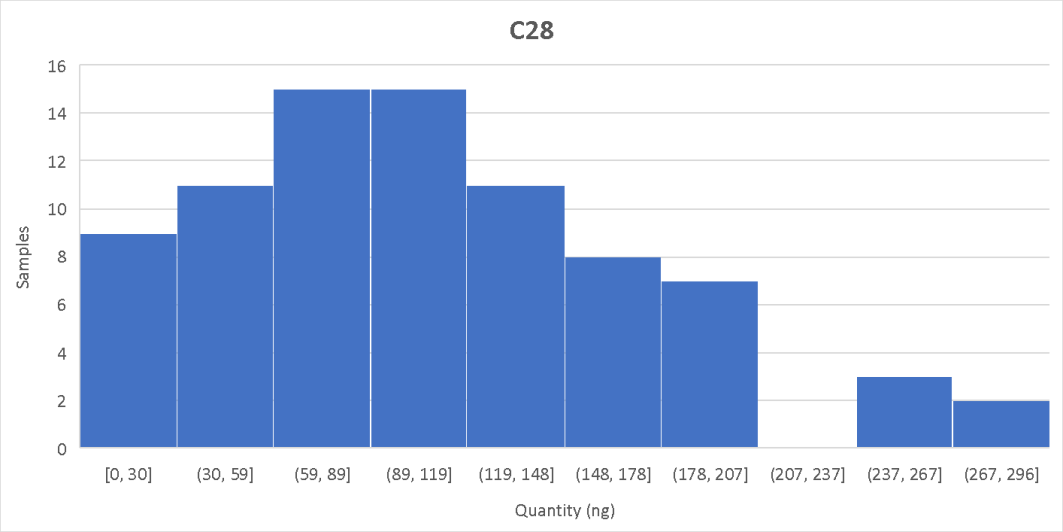
742


7,11DiMe-C27
Samples
70
60
50
40
30
20
10
0
[0, 4]
(4, 7]
(7, 11]
(11, 15]
(15, 19]
(19, 22]
(22, 26]
(26, 30]
(30, 33]
(33, 37]
Quantity (ng)

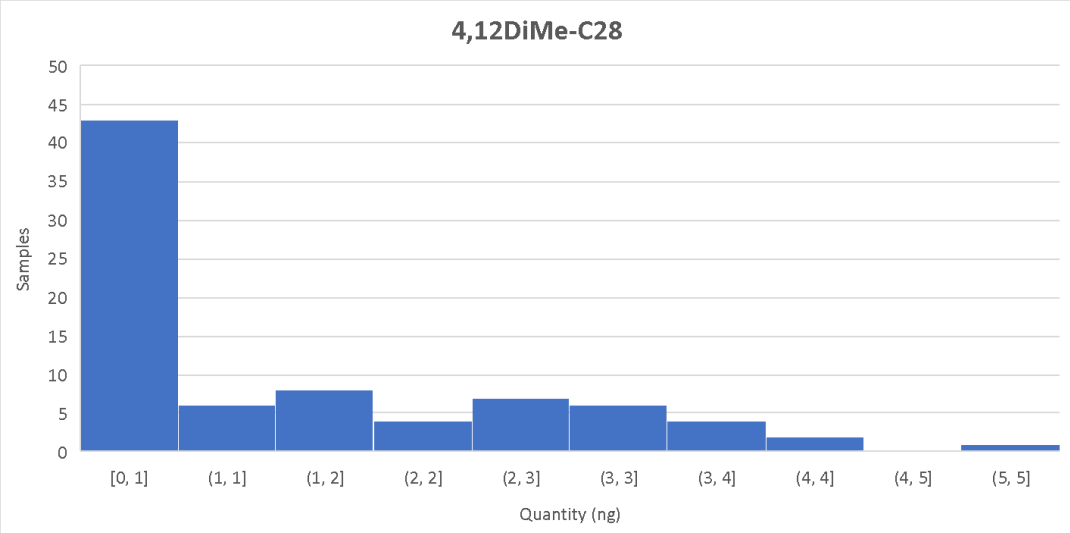
743


11,15DiMe-C27
Samples
30
25
20
15
10
5
0
[0, 4]
(4, 7]
(7, 11]
(11, 14]
(14, 18]
(18, 22]
(22, 25]
(25, 29]
(29, 32]
(32, 36]
Quantity (ng)

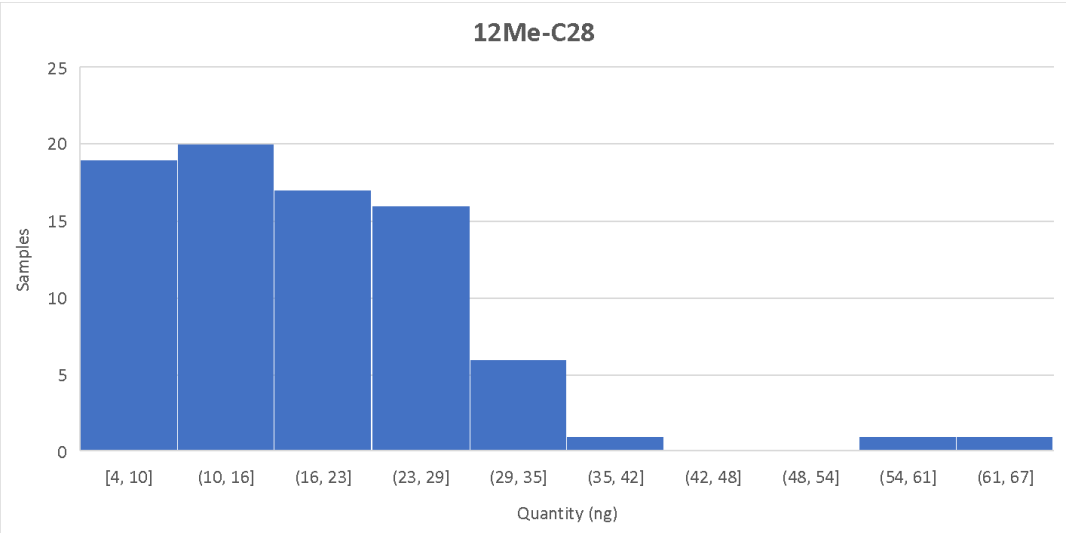
744


C28
Samples
16
14
12
10
8
6
4
2
0
[0, 30]
(30, 59]
(59, 89]
(89, 119]
(119, 148]
(148, 178]
(178, 207]
(207, 237]
(237, 267]
(267, 296]
Quantity (ng)

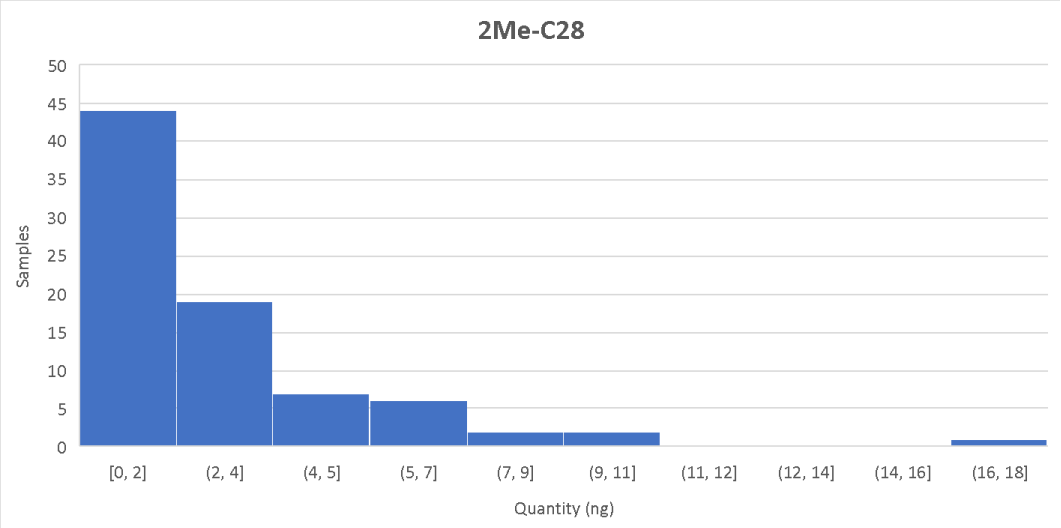
745


4,12DiMe-C28
Samples
50
45
40
35
30
25
20
15
10
5
0
[0, 1]
(1, 1]
(1, 2]
(2, 2]
(2, 3]
(3, 3]
(3, 4]
(4, 4]
(4, 5]
(5, 5]
Quantity (ng)

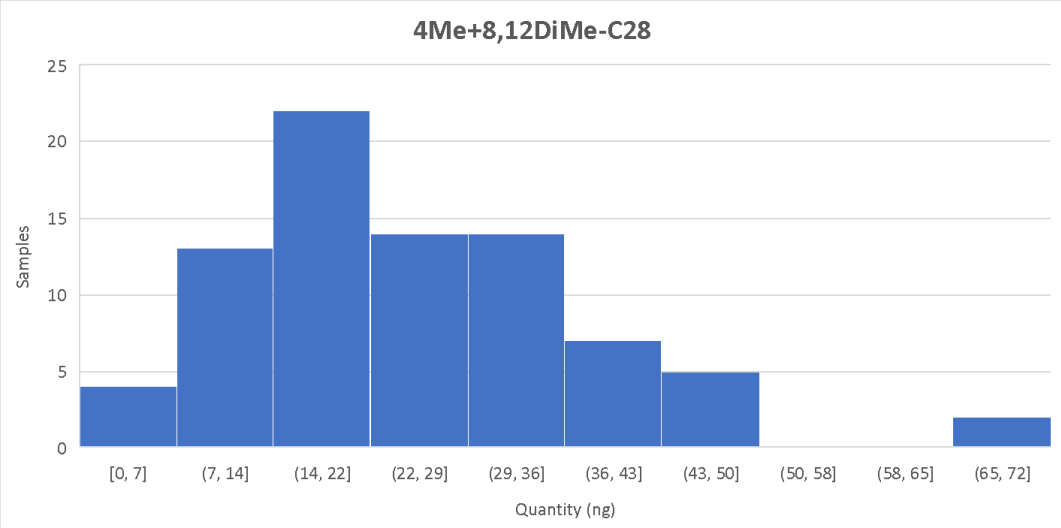
746


12Me-C28
Samples
25
20
15
10
5
0
[4, 10]
(10, 16]
(16, 23]
(23, 29]
(29, 35]
(35, 42]
(42, 48]
(48, 54]
(54, 61]
(61, 67]
Quantity (ng)

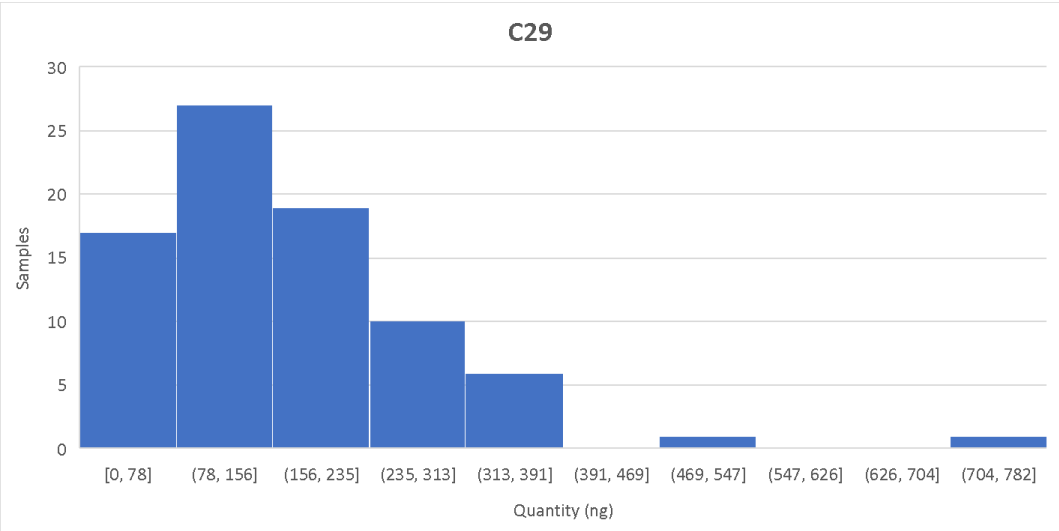
747


2Me-C28
Samples
50
45
40
35
30
25
20
15
10
5
0
[0, 2]
(2, 4]
(4, 5]
(5, 7]
(7, 9]
(9, 11]
(11, 12]
(12, 14]
(14, 16]
(16, 18]
Quantity (ng)

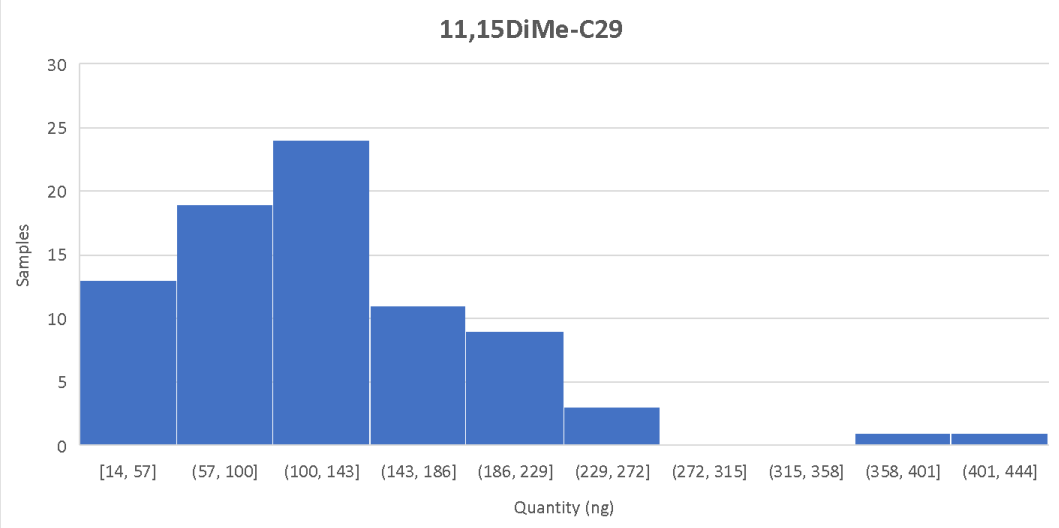
748


4Me+8,12DiMe-C28
Samples
25
20
15
10
5
0
[0, 7]
(7, 14]
(14, 22]
(22, 29]
(29, 36]
(36, 43]
(43, 50]
(50, 58]
(58, 65]
(65, 72]
Quantity (ng)

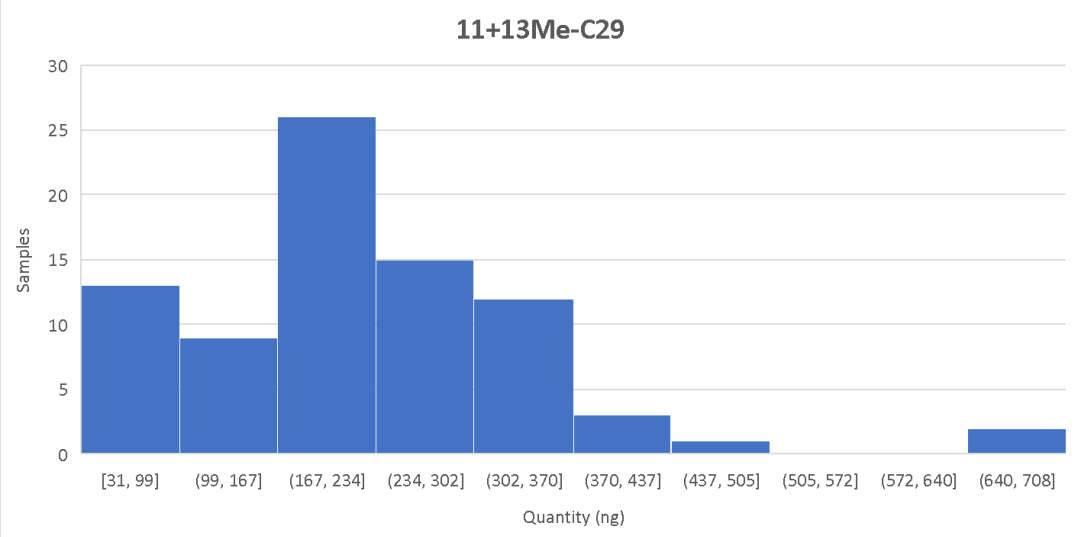
749


C29
Samples
30
25
20
15
10
5
0
[0, 78]
(78, 156]
(156, 235]
(235, 313]
(313, 391]
(391, 469]
(469, 547]
(547, 626]
(626, 704]
(704, 782]
Quantity (ng)

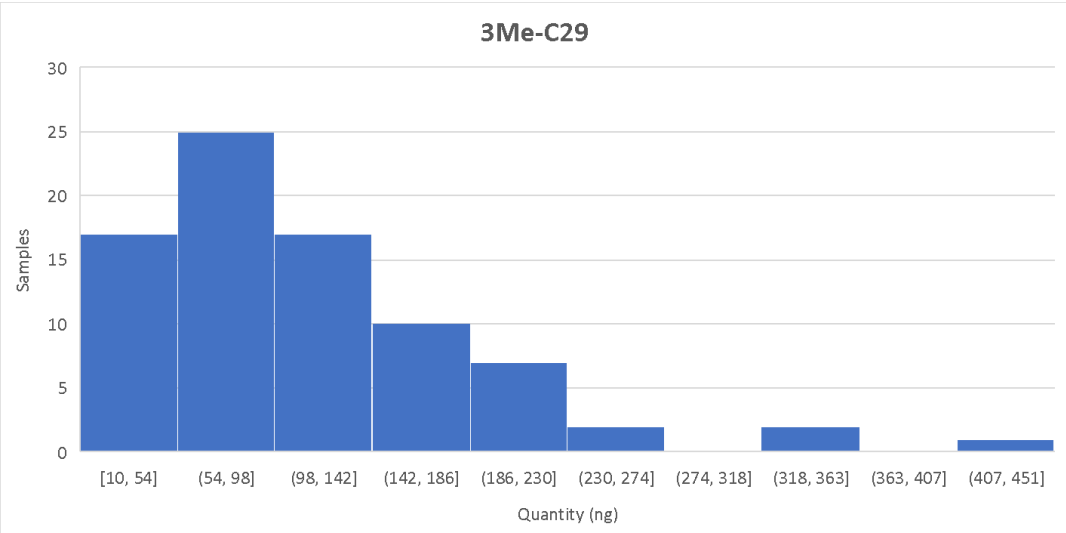
750


11,15DiMe-C29
Samples
30
25
20
15
10
5
0
[14, 57]
(57, 100]
(100, 143]
(143, 186]
(186, 229]
(229, 272]
(272, 315]
(315, 358]
(358, 401]
(401, 444]
Quantity (ng)

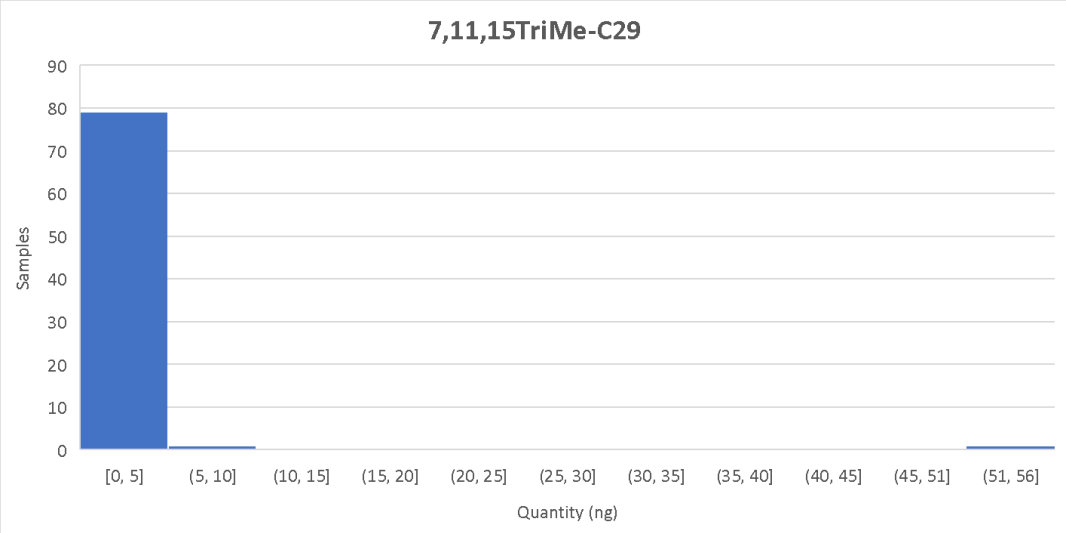
751


11+13Me-C29
Samples
30
25
20
15
10
5
0
[31, 99]
(99, 167]
(167, 234]
(234, 302]
(302, 370]
(370, 437]
(437, 505]
(505, 572]
(572, 640]
(640, 708]
Quantity (ng)

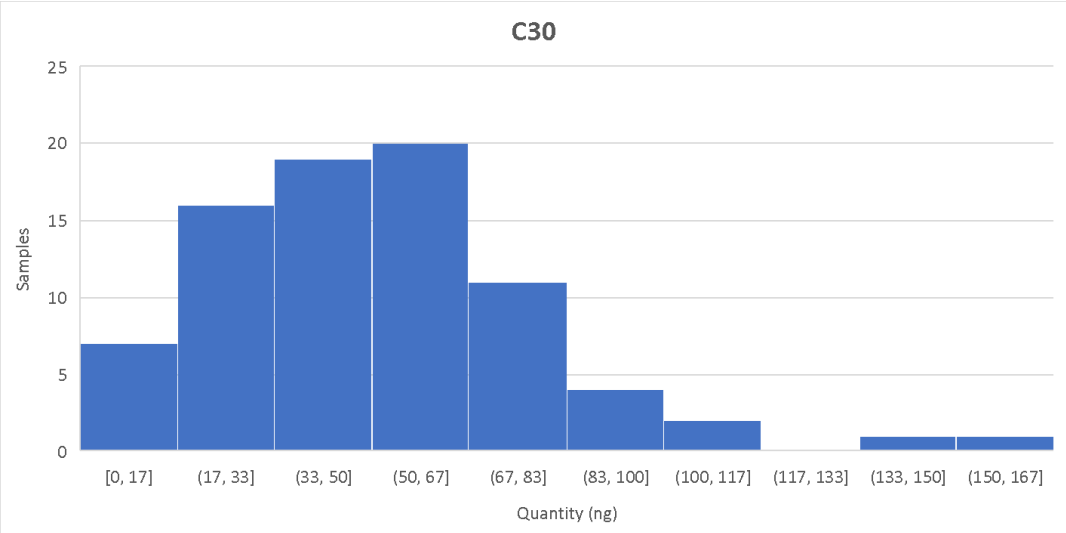
752


3Me-C29
Samples
30
25
20
15
10
5
0
[10, 54]
(54, 98]
(98, 142]
(142, 186]
(186, 230]
(230, 274]
(274, 318]
(318, 363]
(363, 407]
(407, 451]
Quantity (ng)

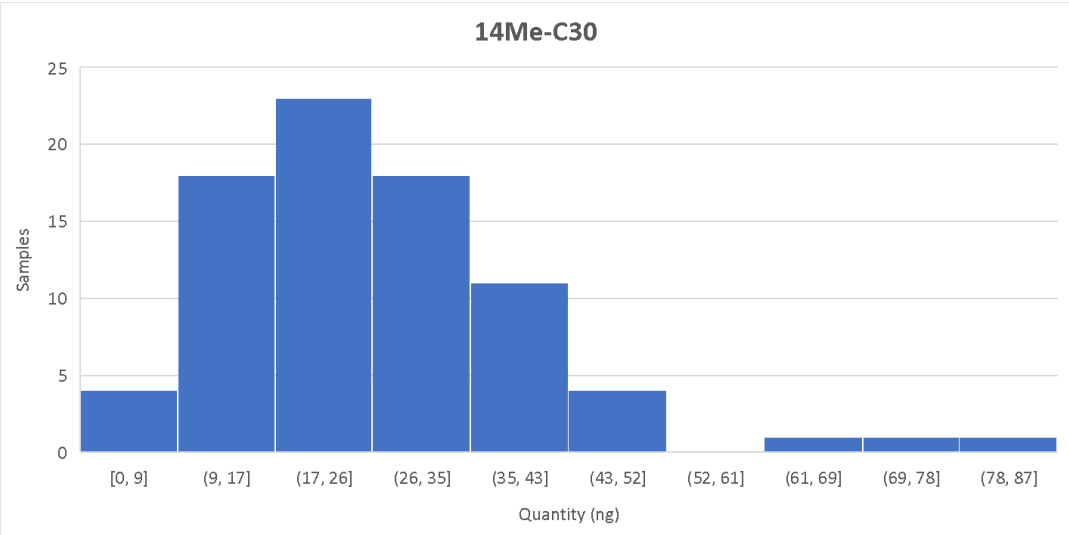
753

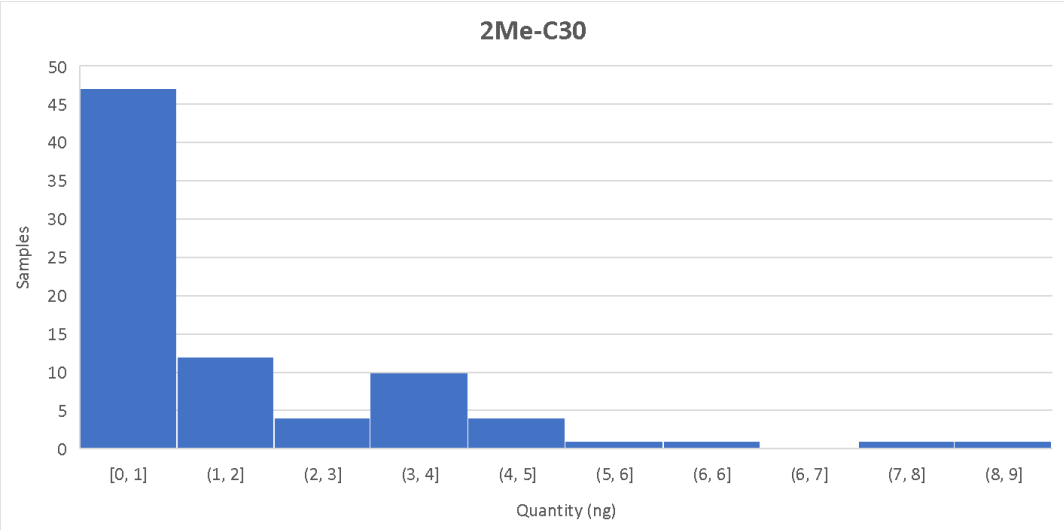

7,11,15TriMe-C29
Samples
90
80
70
60
50
40
30
20
10
0
[0, 5]
(5, 10]
(10, 15]
(15, 20]
(20, 25]
(25, 30]
(30, 35]
(35, 40]
(40, 45]
(45, 51]
(51, 56]
Quantity (ng)

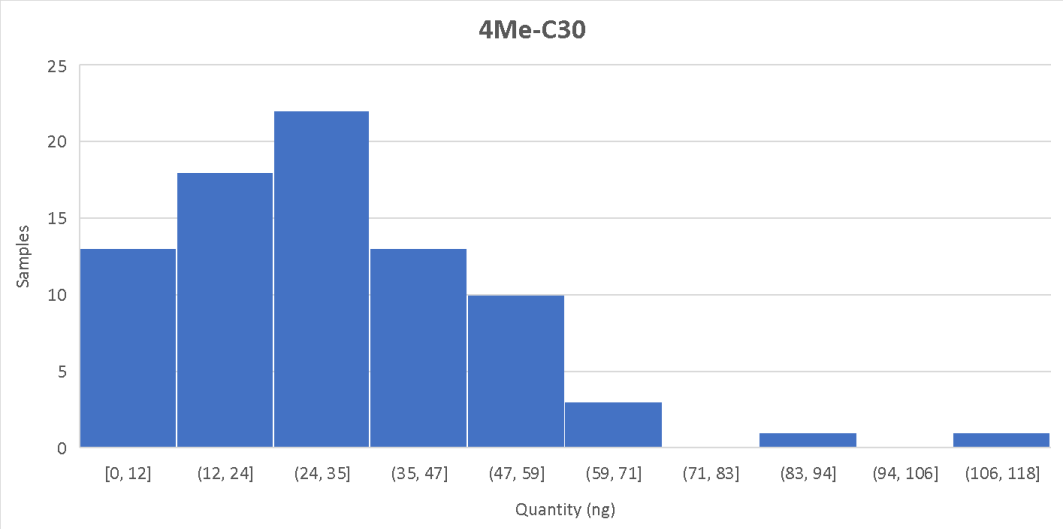
754

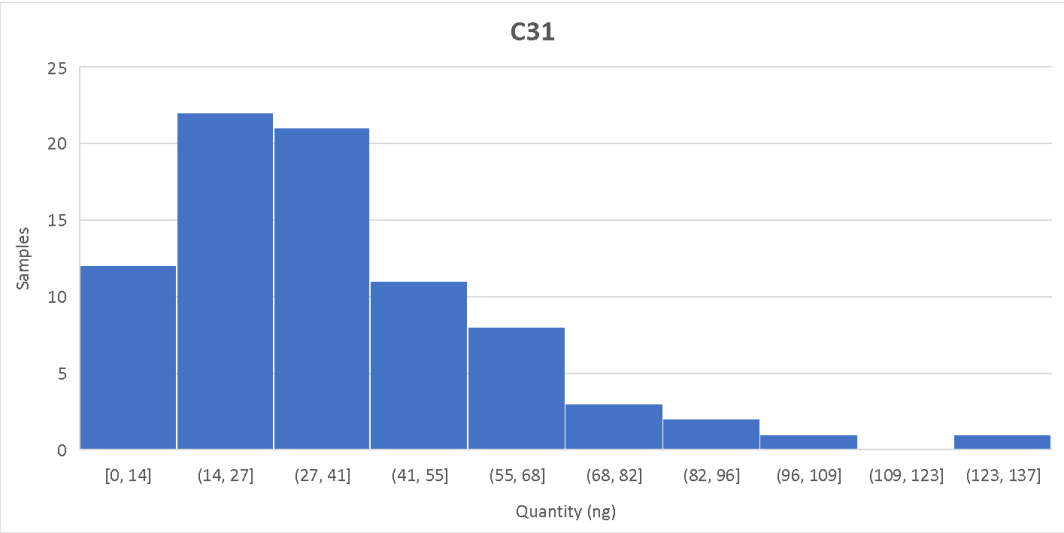

C30
Samples
25
20
15
10
5
0
[0, 17]
(17, 33]
(33, 50]
(50, 67]
(67, 83]
(83, 100]
(100, 117]
(117, 133]
(133, 150]
(150, 167]
Quantity (ng)

755

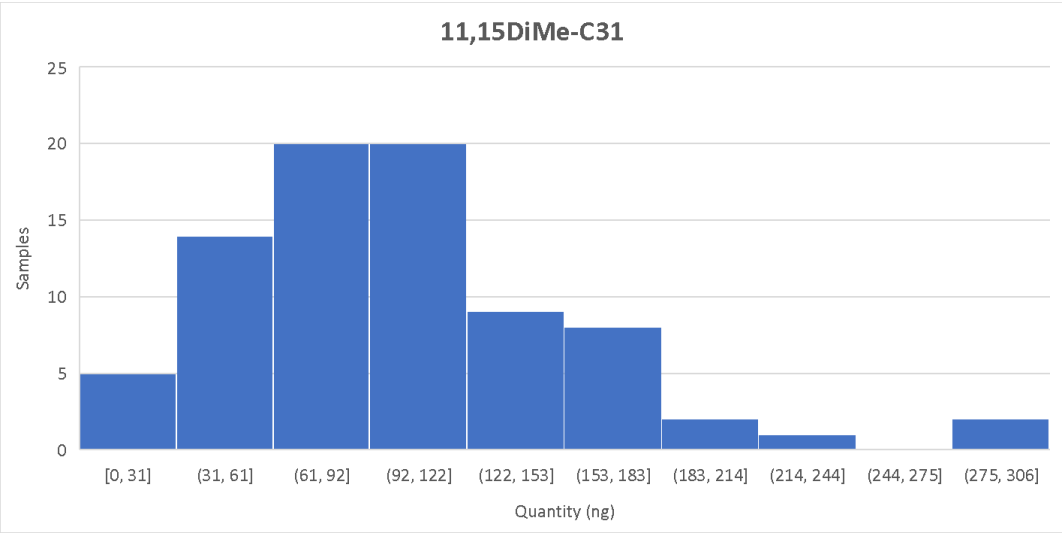

14Me-C30
Samples
25
20
15
10
5
0
[0, 9]
(9, 17]
(17, 26]
(26, 35]
(35, 43]
(43, 52]
(52, 61]
(61, 69]
(69, 78]
(78, 87]
Quantity (ng)


2Me-C30
Samples
50
45
40
35
30
25
20
15
10
5
0
[0, 1]
(1, 2]
(2, 3]
(3, 4]
(4, 5]
(5, 6]
(6, 6]
(6, 7]
(7, 8]
(8, 9]
Quantity (ng)

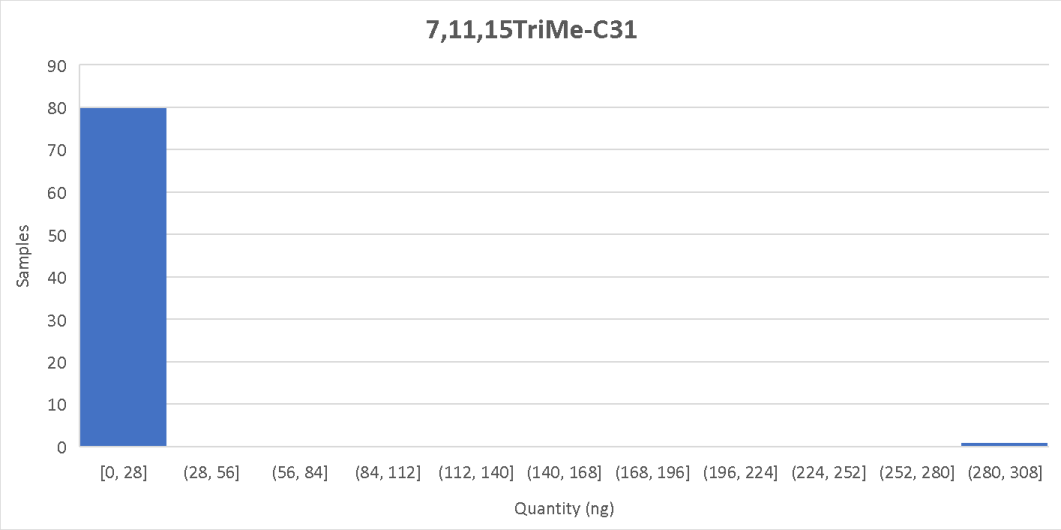

4Me-C30
Samples
25
20
15
10
5
0
[0, 12]
(12, 24]
(24, 35]
(35, 47]
(47, 59]
(59, 71]
(71, 83]
(83, 94]
(94, 106]
(106, 118]
Quantity (ng)


C31
Samples
25
20
15
10
5
0
[0, 14]
(14, 27]
(27, 41]
(41, 55]
(55, 68]
(68, 82]
(82, 96]
(96, 109]
(109, 123]
(123, 137]
Quantity (ng)

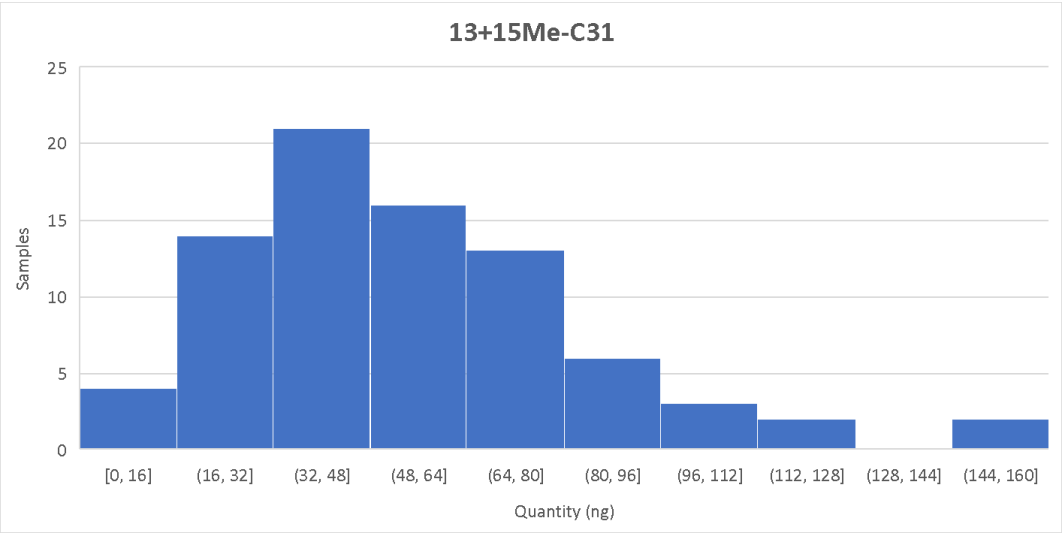
759

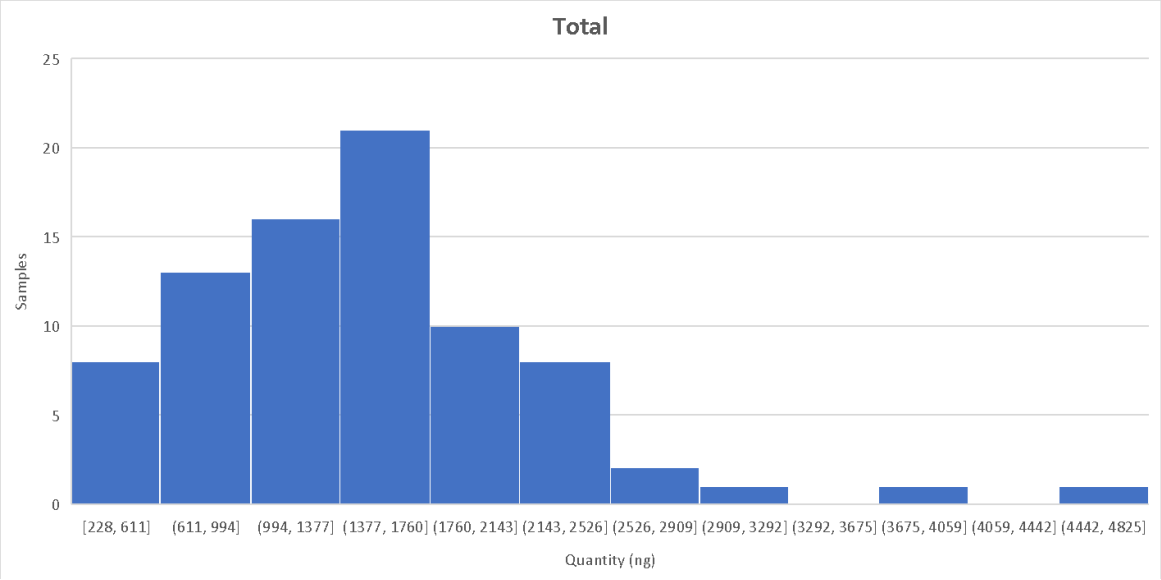

11,15DiMe-C31
Samples
25
20
15
10
5
0
[0, 31]
(31, 61]
(61, 92]
(92, 122]
(122, 153]
(153, 183]
(183, 214]
(214, 244]
(244, 275]
(275, 306]
Quantity (ng)

760

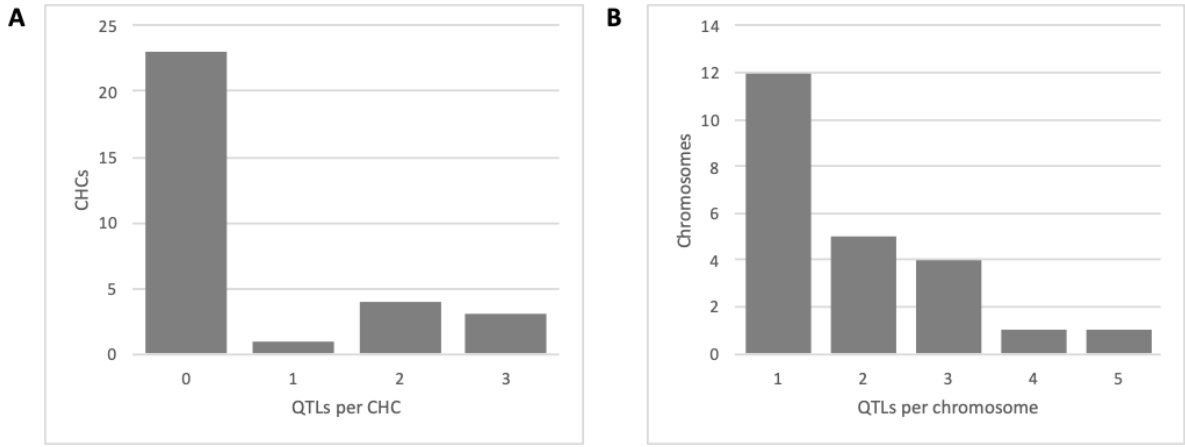

7,11,15TriMe-C31
Samples
90
80
70
60
50
40
30
20
10
0
[0, 28]
(28, 56]
(56, 84]
(84, 112]
(112, 140]
(140, 168]
(168, 196]
(196, 224]
(224, 252]
(252, 280]
(280, 308]
Quantity (ng)

761

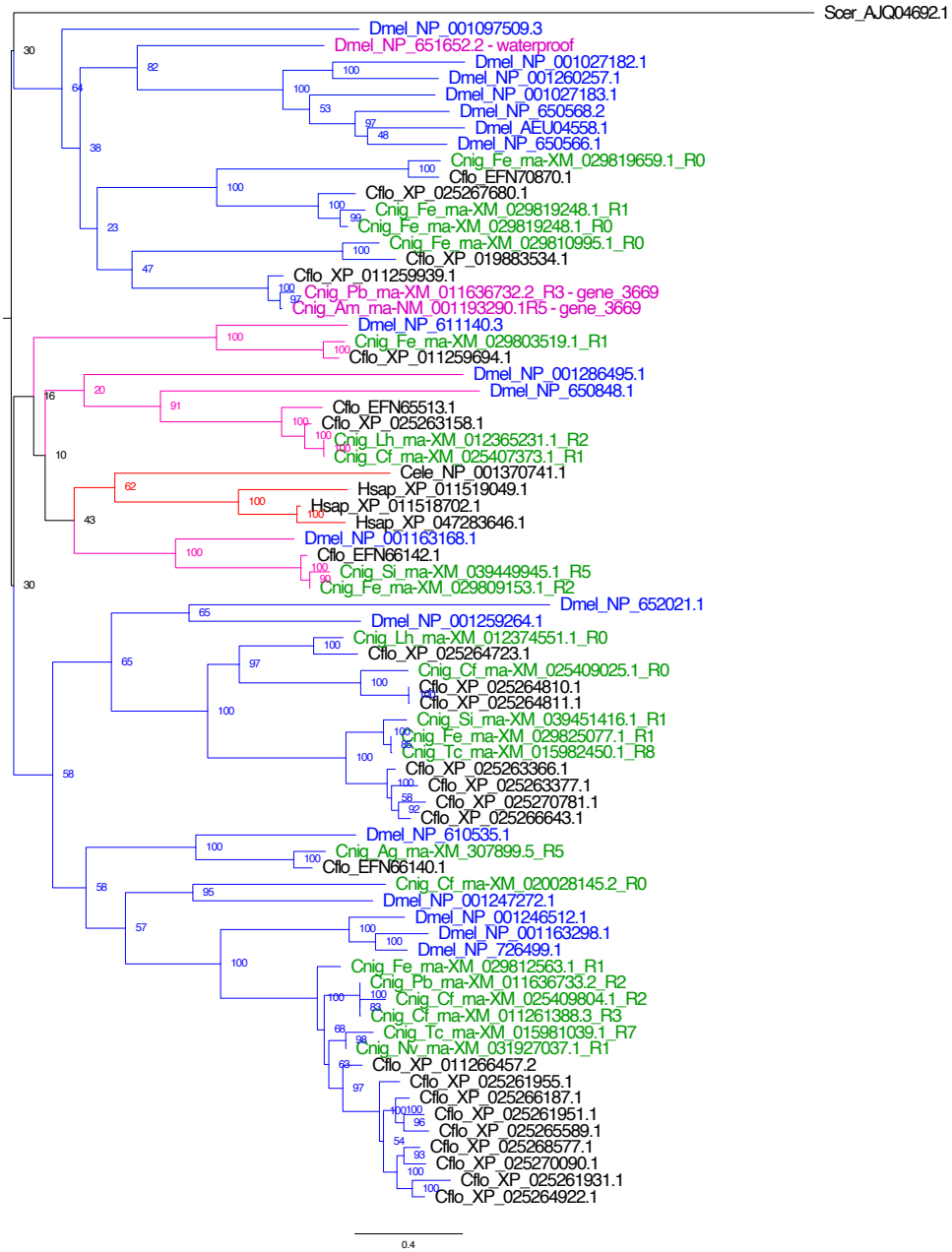




Supplementary Figure S2: Distribution of the number of QTLs per CHC (A) and QTLs per chromosome (B).



Supplementary Figure S3: Fatty acyl-CoA reductase gene tree. Including fly (*Drosophila melanogaster*) and ant (*Cataglyphis niger* and *Camponotus floridanus*) genes. The *Drosophila waterproof* gene and other fly paralogs are the closest homologs in *Drosophila* to *C. niger* gene 3669 in the QTL for C27.



774 **Supplementary Table S1: QTLs underlying variation in CHCs (full table of tested**
775 **candidate QTLs, including non-significant results)**

CHC	LOD	Location (map intervals)	Location (cM)	P.E.V	*p*-value	*q*-value	MIM P.E.V
C25	6.589	Chr6: 1-3-5	0-21.06-43.20	19.41%	0.001	0.069	29.70%
	linked*	Chr6: 19-22-25	76.86-81.62-88.75	linked*	linked*	linked*	
	2.710	Chr8: 15-17-22	58.38-81.91-101.32	10.6%	0.013	0.318	
3Me-C25	2.834	Chr5: 4-8-16	30.89-46.11-66.91	16.8%	0.006	0.207	30.1%
	1.859	Chr1: 50-57-60	386.10-406.87-431.58	13.0%	0.091	1.000	
5Me-C25	2.523	Chr8: 49-50-50	221.43-244.92-244.92	14.9%	0.010	0.301	N/A
C26	4.215	Chr6: 13-16-18	56.40-65.36-67.36	16.2%	0.001	0.069	46.1%
	4.168	Chr17: 17-19-21	60.99-72.57-75.45	16.0%	0.000	0.023	
	3.590	Chr18: 1-1-5	0-0.0-36.50	13.8%	0.001	0.062	
4,12DiMe-C26	2.710	Chr18: 36-39-43	111.63-125.97-130.72	8.7%	0.012	0.318	48.6%
	7.069	Chr3: 1-2-4	0-12.84-37.91	39.9%	0.003	0.125	
	linked*	Chr3: 50-52-56	294.10-313.03-333.81	linked*	linked*	linked*	
4Me-C26	3.349	Chr11: 44-45-48	155.99-172.36-183.23	14.5%	0.003	0.109	35.5%
	2.402	Chr10: 38-43-49	108.13-127.65-154.33	10.4%	0.028	0.503	
	2.507	Chr18: 40-47-52	125.97-143.39-164.19	10.6%	0.018	0.398	
10Me-C26	3.691	Chr9: 45-50-54	190.30-213.93-225.26	16.0%	0.014	0.333	N/A
C27	3.899	Chr17: 25-29-32	87.01-97.50-111.93	17.5%	0.002	0.078	35.4%
	3.981	Chr18: 31-33-37	102.15-108.7-113.54	18.0%	0.000	0.023	
7,11DiMe-C27	3.160	Chr22: 1-1-7	0-0.0-11.61	18.2%	0.004	0.158	N/A
11,15DiMe-C27	2.453	Chr7: 3-4-7	20.58-32.56-44.27	12.4%	0.018	0.398	24.3%
	2.283	Chr22: 1-1-8	0-0.0-13/74	11.8%	0.022	0.447	
3Me-C27	2.250	Chr25: 20-23-28	81.78-106.66-141.64	12.8%	0.018	0.398	N/A
5Me-C27	2.341	Chr14: 1-1-6	0-2.709-19.28	13.6%	0.024	0.484	N/A
11+13Me-C27	2.966	Chr12: 41-43-47	123.83-137.66-145.54	14.8%	0.009	0.283	49.4%
	1.952	Chr10: 5-9-15	27.08-52.66-60.07	10.5%	0.064	0.951	
	1.922	Chr19: 44-46-49	124.68-139.08-143.78	10.5%	0.064	0.951	
	2.760	Chr22: 3-6-10	3.906-11.61-18.61	13.6%	0.011	0.301	
7,11,15TriMe-C27	4.139	Chr15: 43-47-48	122.95-147.06-154.68	15.0%	0.001	0.064	49.4%
	4.810	Chr3: 38-39-42	207.63-216.96-228.07	13.7%	0.000	0.023	
	4.120	Chr6: 41-44-46	191.72-227.46-240.65	12.8%	0.000	0.023	
	4.725	Chr4: 4-8-12	19.14-35.30-45.95	15.9%	0.011	0.301	
	linked*	Chr4: 34-37-46	195.36-209.88-268.45	linked*	linked*	linked*	

C28	2.512	Chr15: 11-15-20	45.45-66.39-71.96	7.1%	0.019	0.405	57.2%
	4.761	Chr10: 34-35-38	97.87-102.49-108.13	14.4%	0.000	0.023	
	6.134	Chr4: 3-7-8	8.909-25.69-33.02	21.9%	0.000	0.023	
	3.910	Chr5: 59-62-64	283.00-300.42-325.22	13.8%	0.002	0.092	
4,12DiMe-C28	4.310	Chr20: 1-4-9	4.967-12.71-28.54	24.9%	0.000	0.023	N/A
	linked*	Chr20: 20-24-30	57.93-73.72-84.35	linked*	linked*	linked*	
2Me-C28	3.001	Chr6: 5-12-16	43.20-53.65-63.35	17.0%	0.010	0.301	N/A
4Me+8,12DiMe-C28	2.119	Chr10: 40-42-49	112.81-127.09-154.33	11.3%	0.045	0.720	21.5%
	1.852	Chr22: 46-51-55	99.76-120.75-130.05	10.8%	0.050	0.785	
12Me-C28	1.86	Chr11: 14-19-23	64.44-83.04-88.70	9.7%	0.032	0.543	20.9%
	2.122	Chr24: 22-27-32	36.70-50.61-58.70	11.2%	0.013	0.318	
C29	3.279	Chr12: 39-43-48	121.89-135.12-153.78	19.0%	0.006	0.208	N/A
11,15DiMe-C29	1.864	Chr16: 15-21-26	70.20-83.69-90.70	10.7%	0.055	0.847	N/A
3Me-C29	2.013	Chr12: 22-29-35	73.42-103.8-116.09	12.0%	0.067	0.978	N/A
11+13Me-C29	4.667	Chr8: 27-31-37	116.69-124.23-143.22	25.2%	0.025	0.484	N/A
	linked*	Chr8: 38-39-42	152.77-158.45-170.69	linked*	linked*	linked*	
7,11,15TriMe-C29	2.766	Chr8: 38-43-45	143.22-173.42-187.41	15.3%	0.013	0.318	28.7%
	2.602	Chr5: 26-29-36	128.50-149.1-163.91	13.4%	0.026	0.484	
C30	2.095	Chr8: 26-31--38	106.23-124.23-143.22	12.3%	0.025	0.484	N/A
2Me-C30	3.823	Chr21: 6-13-17	10.95-26.48-34.17	14.1%	0.001	0.064	46.1%
	4.083	Chr13: 5-9-12	17.33-28.88-31.73	15.6%	0.000	0.023	
	3.838	Chr4: 1-2-4	0-6.219-17.23	16.4%	0.002	0.092	
4Me-C30	2.116	Chr12: 10-14-19	36.87-55.90-64.78	12.5%	0.043	0.716	N/A
14Me-C30	2.027	Chr6: 25-27-30	88.74-113.07-127.81	12.0%	0.078	1.000	25.5%
	2.398	Chr18: 27-36-39	87.11-113.54-122.12	13.9%	0.026	0.484	
C31	1.813	Chr3: 12-18-21	92.83-142.88-166.87	13.0%	0.150	1.000	26.1%
	2.386	Chr21: 46-52-52	112.22-135.98-139.79	13.6%	0.005	0.189	
11,15DiMe-C31	2.022	Chr23: 7-15-23	14.06-36.32-556.53	13.1%	0.032	0.543	N/A
13+15Me-C31	4.837	Chr18: 32-35-38	105.86-111.63-120.25	25.4%	0.000	0.023	N/A
Total	2.141	Chr14: 26-30-39	77.33-97.10-115.14	8.0%	0.045	0.720	44.0%
	3.886	Chr15: 26-29-31	89.34-98.91-102.52	15.1%	0.001	0.064	
	3.065	Chr3: 29-32-36	187.59-194.33-205.79	11.8%	0.011	0.305	
	2.413	Chr7: 13-17-21	65.37-76.84-83.48	9.1%	0.031	0.543	

CHC = cuticular hydrocarbon; LOD = log-of-odds score; Location = location of confidence interval start and end, and of the maximal LOD score; *q*-value = FDR corrected *p*-value; PEV = percent explained variation; MIM = multi-interval model summarizing all QTLs for a given CHC over different chromosomes.

779 * Two QTLs for the same CHC on the same chromosome evaluated using the linked-QTL model of *MultiQTL*.
780 Therefore, the LOD, P.E.V., *p*-value, and *q*-value are for both QTLs together.
781