Genetic and genomic architecture of species-specific cuticular hydrocarbon variation in parasitoid wasps

Jan Buellesbach[*], Henrietta Holze[*], Lukas Schrader[*], Jürgen Liebig[†], Thomas Schmitt[‡], Juergen Gadau[*] & Oliver Niehuis[§]

[*] Molecular Evolution and Sociobiology Group, Institute for Evolution & Biodiversity, University of Münster, Hüfferstr. 1, DE-48149 Münster, Germany

[†] School of Life Sciences, Arizona State University, P.O. Box 874701, Tempe, Arizona 85287-4501, USA

[‡] Department of Animal Ecology and Tropical Biology, Biocenter, University of Würzburg, Am Hubland, D-97074 Würzburg, Germany

[§]Department of Evolutionary Biology and Ecology, Institute of Biology I (Zoology), Albert Ludwig University of Freiburg, Hauptstr. 1, 79104 Freiburg, Germany

Running title: Wasp cuticular hydrocarbon genetics

Keywords: comparative genomics, gene mapping, ortholog inference, quantitative trait loci, chemical ecology, entomology, *Drosophila*, *Nasonia*, Hymenoptera, Pteromalidae

Corresponding author: Jan Buellesbach, Institute for Evolution & Biodiversity (Room 221), Hüfferstr. 1, DE-48149 Münster, Germany, Phone: +49 (0) 251-83-21637, Email: buellesb@uni-muenster.de

ABSTRACT

Cuticular hydrocarbons (CHCs) serve two fundamental functions in insects: protection against desiccation and chemical signaling. CHC profiles can consist of dozens of different compounds and are considered a prime example for a complex trait. How the interaction of genes shapes CHC profiles, which are essential for insect survival, adaptation, and reproductive success, is still poorly understood. Here we investigate the genetic and genomic basis of CHC biosynthesis and variation in parasitoid wasps of the genus *Nasonia*. Taking advantage of the wasps' haplo-diploid sex determination and cross-species fertility, we mapped 91 quantitative trait loci (QTL) explaining variation of a total of 43 CHCs in F_2 hybrid males from interspecific crosses between three *Nasonia* species. To identify candidate genes, we localized orthologs of CHC biosynthesis-related genes in the *Nasonia* genomes. By doing so, we discovered multiple genomic regions where the location of QTL coincides with the location of CHC biosynthesis-related candidate genes. Most conspicuously, on a region on chromosome 1 close to the centromere, multiple CHC biosynthesis-related candidate genes co-localize with several QTL explaining variation in methyl-branched alkanes. The genetic underpinnings behind this compound class are not well understood so far, despite their high potential for encoding chemical information as well as their prevalence in both *Nasonia* CHC profiles and many other Hymenoptera. Our study considerably extends our knowledge on the so far little-known genetic and genomic architecture governing biosynthesis and variation of this fundamental compound class, establishing a model for methyl-branched alkane genetics in the Hymenoptera in general.

INTRODUCTION

Understanding the genetic basis of quantitative phenotypic traits remains one of the central challenges in population genetics and evolutionary biology despite several advances in this field during the last decades (Yang et al. 2007; Mackay et al. 2009; Hartl 2020). Genes and regulatory nucleotide sequences governing natural variation in a set of phenotypic traits can be mapped (and potentially identified) in a genome by searching for phenotypic covariation with polymorphic genetic markers. This technique is commonly referred to as quantitative trait locus (QTL) mapping (Lander and Botstein 1989; Haley and Knott 1992; Mackay 2001a). Evolutionary genetic studies addressing variation and divergence of traits between closely related species with varying degrees of phylogenetic distance between them particularly benefit from QTL mapping techniques (Orr 2001; Slate 2005; Hartl 2020). Two components are of paramount importance for these kinds of studies: first, the availability of genetically tractable and crossable species with comparably short generation times and, ideally, established phylogenetic relationships (Mackay 2001b; Orr 2001; Mackay et al. 2009). Second, polygenic traits of interest that differ between the crossed species (Erickson et al. 2004; Slate 2005). Insects have proven to be particularly useful for QTL studies, since they combine various beneficial attributes: a multitude of quantifiable phenotypes, short generation times, large offspring numbers and, most commonly, relatively small genomes (Hoy 2005; Hunter and Kole 2008; Thomas et al. 2020). Cuticular hydrocarbons (CHCs), lipids constituting a major part of the waxy layer covering the insects' epicuticle, are particular promising traits to dissect genetically via QTL mapping, as they are comparatively easily quantifiable, polygenic, and phenotypically as well as genotypically traceable (Chung and Carroll 2015; Blomquist and Ginzel 2021; Holze et al. 2021).

92

CHCs are long-chained molecules whose primary function is to protect terrestrial
94 insects from desiccation (Ramsay 1935; Lockey 1980; Howard and Blomquist 1982, Fig. 1). As versatile semiochemicals, CHCs have also been demonstrated to function in a wide array of
96 chemical communication systems integral to insect survival, reproduction, and adaptation (Carlson et al. 1971; Blomquist and Bagnères 2010; Blomquist and Ginzel 2021). Differences
98 between CHC profiles arise from presence, absence, and abundance of individual CHCs, which can differ in their chain length as well as in the presence and position of double bonds and
100 methyl groups (Lockey 1988; Howard and Blomquist 2005). The most commonly occurring compound classes are saturated straight-chain alkanes (*n*-alkanes), unsaturated straight-chain
102 alkenes (*n*-alkenes) and alkadienes, and methyl-branched alkanes (Blomquist et al. 1987; Blomquist and Bagnères 2010). Despite considerable diversity of CHC profiles across insects,
104 to the current stage of knowledge, the pathway of CHC biosynthesis appears to be mostly conserved (Blomquist and Ginzel 2021; Holze et al. 2021). Briefly, the biosynthetic pathway
106 consists of the elongation of fatty-acyl-Coenzyme A units to produce very long-chain fatty acids (VLCFs) that are subsequently converted to hydrocarbons by subducting the carboxyl
108 group (Nelson and Blomquist 1995; Howard and Blomquist 2005; Blomquist and Bagnères 2010, Fig. 1).

110

Most of what we currently know about the genetics governing CHC biosynthesis and
112 CHC profile diversity is based on studies on species of the Dipteran genus *Drosophila* (e.g. Jallon 1984; Coyne and Oyama 1995; Ferveur 2005). This taxonomic bias has so far impeded a
114 more holistic view on the genetic underpinnings of this fundamental process across insect orders (Blomquist and Bagnères 2010; Holze et al. 2021). Particularly little is known about the

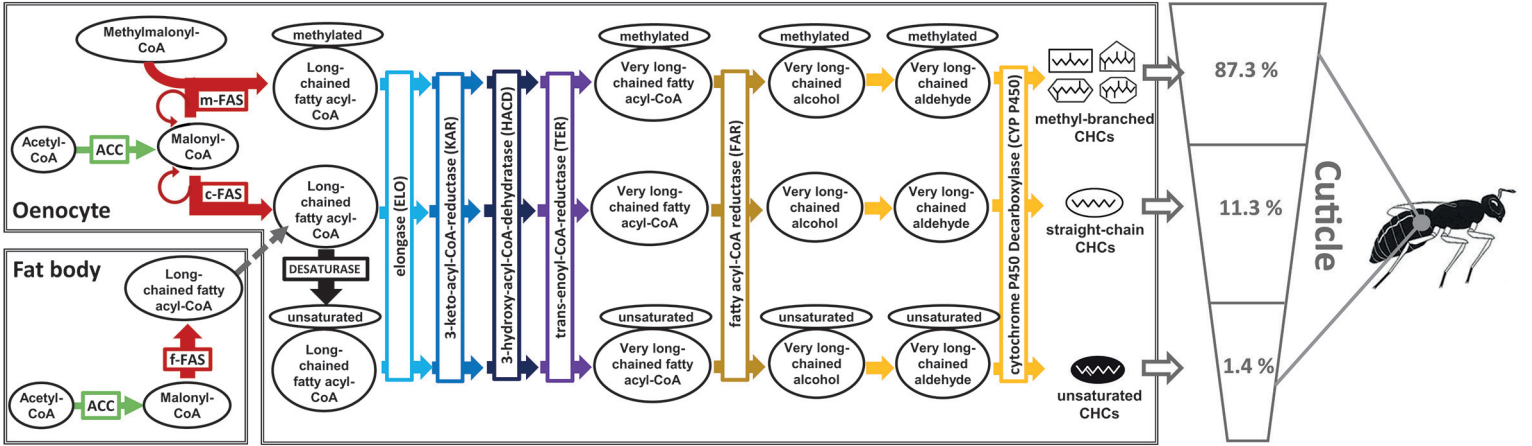
Oenocyte
Methylmalonyl-CoA
m-FAS
Acetyl-CoA
ACC
Malonyl-CoA
c-FAS
Fat body
Long-chained fatty acyl-CoA
f-FAS
Acetyl-CoA
ACC
Malonyl-CoA
methylated
Long-chained fatty acyl-CoA
Long-chained fatty acyl-CoA
DESATURASE
unsaturated
Long-chained fatty acyl-CoA
elongase (ELO)
3-keto-acyl-CoA-reductase (KAR)
3-hydroxy-acyl-CoA-dehydratase (HACD)
trans-enoyl-CoA-reductase (TER)
methylated
Very long-chained fatty acyl-CoA
Very long-chained fatty acyl-CoA
unsaturated
Very long-chained fatty acyl-CoA
fatty acyl-CoA reductase (FAR)
methylated
Very long-chained alcohol
Very long-chained alcohol
unsaturated
Very long-chained alcohol
methylated
Very long-chained aldehyde
Very long-chained aldehyde
unsaturated
Very long-chained aldehyde
cytochrome P450 Decarboxylase (CYP P450)
methyl-branched CHCs
straight-chain CHCs
unsaturated CHCs
87.3 %
11.3 %
1.4 %
Cuticle

Figure 1: Simplified overview of CHC biosynthesis, with the main enzymes catalyzing the intermediate reactions highlighted (compare to Fig. 3-5 for the localization of the corresponding genes). The pathway branches at different stages, eventually synthesizing the different CHC compound classes that are then transported to the cuticle. The main CHC compound classes are methyl-branched alkanes (mono-, di-, tri-, and tetra-), straight-chain alkanes, and unsaturated alkanes. Acetyl-CoA as the initial reactant of CHC biosynthesis is mainly provided by the citric acid cycle. Respective percentages of the individual compound classes found on the cuticle of *Nasonia* males (averaged over our three study species *N. vitripennis*, *N. longicornis* and *N. giraulti*) are indicated as well. Schematic insect drawing depicting an *N. vitripennis* male adapted from Ward's science, Rochester, NY, USA, 2002. Abbreviations: CoA: Coenzyme A, ACC: Acetyl-CoA carboxylase, FAS: Fatty acid synthase (m: microsomal, c: cytosolic, f: fat body associated). Adapted from Holze et al (2021).

genetic factors governing the differentiation and variation of methyl-branched alkanes, a dominant compound class in many insect taxa with high potential for encoding chemical information (Niehuis et al. 2011; Chung and Carroll 2015; Holze et al. 2021). In contrast to *Drosophila*, where methyl-branched alkanes only comprise a small fraction of their CHC profiles, this particular compound class dominates the CHC profiles in most species of the insect order Hymenoptera (Martin and Drijfhout 2009; Kather and Martin 2015; Sprenger and Menzel 2020). However, no single gene with a demonstrated impact on methyl-branched CHC biosynthesis and variation has been unambiguously identified in Hymenoptera so far. To address this knowledge gap, suitable model organisms are required to provide a basis for studying the genetic background of CHC production, variation, and divergence.

The parasitoid jewel wasp genus *Nasonia* (Hymenoptera: Pteromalidae) has emerged as a model system well-suited to study quantitative as well as polygenic phenotypic traits in the Hymenoptera. It combines ease of maintenance and handling, the possibility to generate hybrids between its four described species, haplo-diploid genetics, and a growing genetic tool kit (Gadau et al. 2008; Werren and Loehlin 2009; Werren et al. 2010). The *Nasonia* species complex consists of four species, *N. vitripennis*, *N. longicornis*, *N. giraulti*, and *N. oneida* (Darling and Werren 1990; Raychoudhury et al. 2010). Natural populations of *Nasonia* are all infected with different strains of intracellular *Wolbachia* bacteria, leading to bidirectional incompatibility in most interspecific crosses (Breeuwer and Werren 1990; Bordenstein and Werren 1998; Bordenstein et al. 2001). However, the *Wolbachia* infections can be cured (Richardson et al. 1987), rendering hybrids between all four species readily obtainable (Breeuwer and Werren 1990; Bordenstein and Werren 1998; Raychoudhury et al. 2010). This has proven to be particularly useful for QTL mapping of interspecific differences in a variety of

140 phenotypic traits (Gadau et al. 2002; Gadau et al. 2008; Niehuis et al. 2008). An exemplary study comparing CHC variation through QTL mapping between two *Nasonia* species,
142 *N. vitripennis* and *N. giraulti*, shed some first light on the genetic basis of *Nasonia* CHC profile differences (Niehuis et al. 2011). A study on CHC variation between all so far described
144 *Nasonia* species established clearly distinguishable species- and sex-specific CHC blends (Buellesbach et al. 2013).

146

Here we dissect the genomic and genetic architecture of CHC variation in hybrid cross-
148 comparisons between three *Nasonia* species. We present results from conducting QTL analyses on CHC variation in recombinant F_2 hybrid males obtained by crossing *N. vitripennis*,
150 *N. longicornis*, and *N. giraulti,* and from mapping CHC biosynthesis candidate genes. We additionally screened for selection signatures considering candidate gene orthologs from two
152 more distantly related parasitoid wasp species within the family Pteromalidae: *Muscidifurax raptorellus* and *Trichomalopsis sarcophagae* (Fig. 2). Finally, we discuss how CHC profile
154 divergence between the three *Nasonia* species can be explained phenotypically when considering the newly gained knowledge on the genetic architecture of CHC profile differences
156 between them.

158

160

162

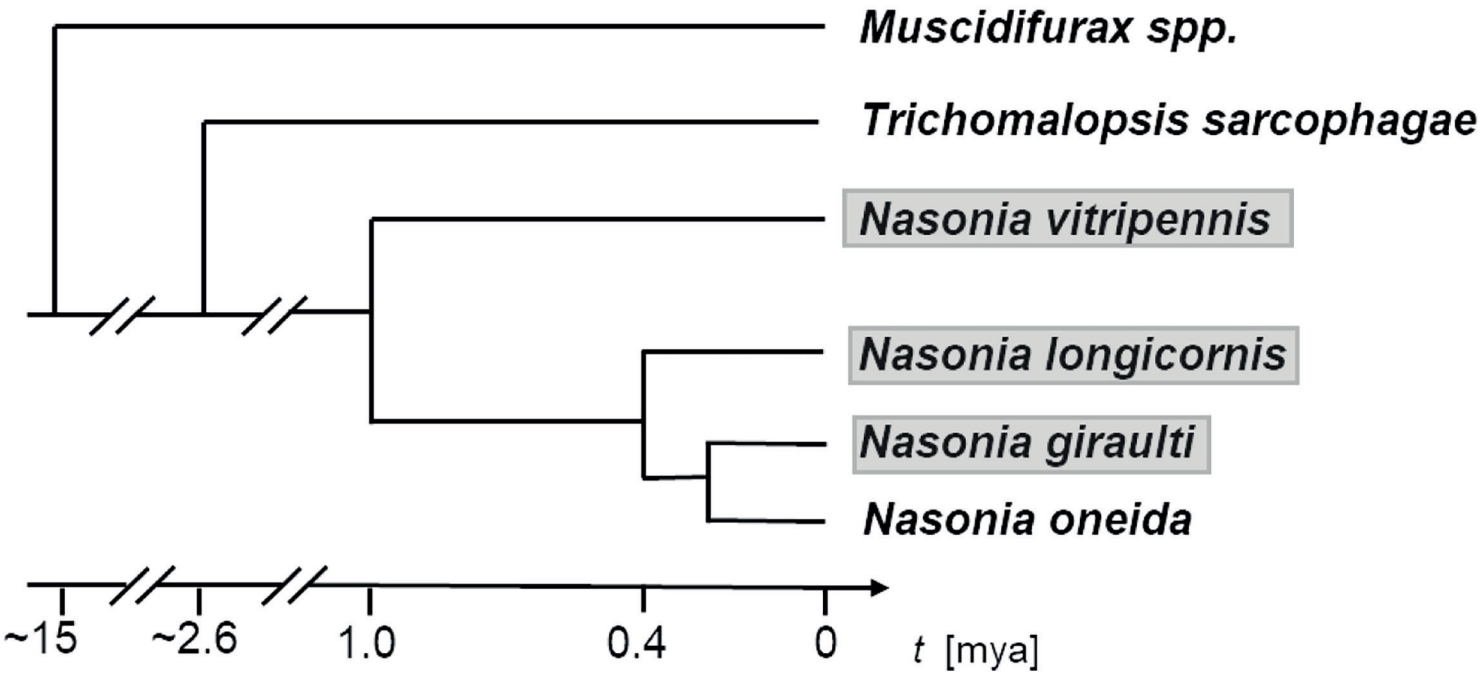
Muscidifurax spp.
Trichomalopsis sarcophagae
Nasonia vitripennis
Nasonia longicornis
Nasonia giraulti
Nasonia oneida
~15
~2.6
1.0
0.4
0
t [mya]

Figure 2: Phylogenetic relationship of the *Nasonia* species complex rooted by representatives of its two most closely related genera, *Trichomalopsis* and *Muscidifurax*, based on the D2 expansion region of their 28S rDNA sequences (Campbell et al. 2000) and mitochondrial DNA sequences (Werren et al. 2010). The three *Nasonia* species used in the current study are highlighted in grey.

MATERIALS AND METHODS

Nasonia strains and crossing experiments

We used the standard laboratory strains AsymCX, RV2x(U), and IV7(U) of the species *Nasonia vitripennis*, *N. giraulti*, and *N. longicornis*, respectively, to conduct the cross experiments (Breeuwer and Werren 1995; Werren et al. 2010). These strains are derived from wild-type strains and have been antibiotically cured from their *Wolbachia* infection, enabling them to hybridize under laboratory conditions. All strains were maintained in an incubator at 25 °C under permanent light and were provided pupae of the flesh fly *Sarcophaga bullata*. Cross experiments were conducted as described by Niehuis et al. (2008). We generated two types of F1 hybrid females: 1) by crossing virgin *N. vitripennis* females with *N. longicornis* males and 2) by crossing virgin *N. giraulti* females with *N. longicornis* males. All F1 hybrid females were kept unmated and thus generated recombinant (i.e., each genome constitutes a unique combinations of the parental genomes) haploid F2 males. These F2 recombinant hybrid males were collected 36 h after they had eclosed and were then freeze-killed and stored at -80 °C before characterizing their genotypes and CHC phenotypes.

CHC profile analyses

We analyzed the CHC profiles of 100 *N. vitripennis* x *N. longicornis* (LV) and 101 *N. longicornis* x *N. giraulti* (LG) recombinant hybrid F2 males. We additionally characterized the CHC profiles of twelve *N. giraulti*, eleven *N. longicornis*, and 32 *N. vitripennis* males of the investigated strains for comparison. CHCs were extracted by submersing individual wasps for 10 min in 10

186 µl n-hexane (≥ 99.0%; Sigma-Aldrich, St. Louis, MO, USA) using 1-ml glass vials equipped with a 0.1-ml micro insert (Alltech, Deerfield, IL, USA). The CHC extracts were subsequently
188 transferred to new vials and concentrated under a stream of nitrogen to approximately 1 µl total volume. The CHC extracts were then injected into a gas chromatograph coupled with a
190 mass spectrometer (GC: 6890N; MS: 5975C; Agilent, Santa Clara, CA, USA) for analysis. The GC-MS was operated with in the splitless mode with an injector temperature of 250 °C.
192 Separation of compounds was performed on a J&W DB-5MS fused silica capillary column (30 m × 0.25 mm ID; df = 0.25 µm; Agilent, Santa Clara, CA, USA) and applying the following
194 temperature program: 60 °C start temperature, temperature increase by 40 °C per min to 200 °C, followed by an increase of 5 °C per min to 320 °C. Helium with a constant flow of 1 ml per
196 min was used as carrier gas. The MS-quad temperature was 150 °C, the MS-source temperature was 230 °C, and the MSD-transfer line temperature was 315 °C. The obtained
198 chromatograms were analyzed with the software Enhanced Chemstation E.01.00.237 (Agilent, Santa Clara, CA, USA). Peak identification was carried out automatically with the initial area
200 reject parameter set to 0, the initial peak width set to 0.043, and the initial threshold set to 14. Shoulder detection was turned off. CHC compounds were identified according to their
202 retention indices, diagnostic ions, and mass spectra (Carlson et al. 1998; Niehuis et al. 2011; Buellesbach et al. 2013).

204

206 ***DNA extraction and QTL analysis***

We extracted genomic DNA of all wasps after extraction of their CHCs. DNA extraction was
208 done as outlined by Niehuis et al. (2007). Genotype data were collected for 29 length-polymorphic markers with known positions in the *Nasonia* nuclear genome (Tab. S1, Niehuis

210 et al. 2011). The species-specific length of the markers was assessed by separating PCR-generated amplicons of them on a denaturing polyacrylamide gel following the procedure
212 given by Niehuis et al. (2008). We used the following PCR temperature program for obtaining the amplicon: 5 min. at 95 °C, followed by 30 cycles of 1 min. at 95 °C, 1 min. at 55 °C, and 1
214 min. at 72 °C, and followed by 10 min. at 72 °C. PCR oligonucleotide primer sequences with their respective amplicon lengths are listed in Tab. S1.

216 Quantitative trait loci (QTL) analyses were conducted with the R/qtl package version 1.14-2 (Broman et al. 2003; Broman and Sen 2009) on a 64-bit-build of R version 2.10.1 (R
218 Development Core Team 2018). Significant phenotypic differences in the relative abundances of CHC compounds were assessed with Bonferroni-corrected Wilcoxon rank-sum tests with
220 continuity correction (Hollander and Wolfe 1999; Zar 1999). Each CHC compound was subjected to a one-dimensional one-QTL and to a two-dimensional two-QTL scan using Haley-
222 Knott regression (Haley and Knott 1992) with an assumed genotypic error probability of 0.001 and a step width of 1 cM. We additionally applied a multiple QTL model to each CHC trait with
224 forward/backward model selection as implemented in R/qtl. The maximum number of QTL for each trait was set to 10, with QTL positions refined after each step of the forward and
226 backward selection. Penalties for model selection were set based on the permutation from the two-dimensional two-QTL genome scan of the respective CHC trait with a significance level
228 of 0.01 (Manichaikul et al. 2009). Significance thresholds for QTL presence were estimated from 10,000 permutations of each respective phenotypic CHC abundance. Since our QTL
230 mapping was based on the *Nasonia* linkage map inferred by (Niehuis et al. 2010) and on genome assembly 1.0 (Nvit 1.0, Werren et al. 2010), we transferred marker positions and QTL
232 to the most recent high-resolution linkage map of *Nasonia* (Nvit 2.1, Desjardins et al. 2013) where we performed the mapping of candidate CHC biosynthesis genes (Tab. S1, see below).

234 Though the position of several markers had slightly shifted on the new linkage map, their respective initial linkage groups remained the same, enabling the re-assessment of the QTL
236 positions by their relative position between the two closest markers. A fixed confidence interval of 20 cM was chosen for each QTL, since this distance was equivalent to the average
238 distance between markers.

240

Search for candidate CHC biosynthesis genes

242 Thirty-eight well-characterized candidate genes with a demonstrated impact on CHC variation via targeted knockdown studies were selected from *Drosophila melanogaster* (Tab. S2, see
244 also Holze et al. 2021) via FlyBase, Version FB2019_02 (http://flybase.org/, Gramates et al. 2017). To screen for orthologs of these candidate genes in the *Nasonia* Nvit 2.0 reference
246 genome, three complementary methods were used: first, the amino-acid sequences, including all possible isoforms encoded by *D. melanogaster* CHC biosynthesis genes of interest (GOI),
248 were searched against the *N. vitripennis* proteome available on NCBI (*N. vitripennis*, https://www.ncbi.nlm.nih.gov/) using the BLAST+ software suite (version 2.8.1, Camacho et
250 al. 2009). Amino-acid sequences with an e-value smaller than 10^{-10} were then again searched reciprocally against the *D. melanogaster* proteome. Reciprocal best hits were considered as
252 orthologs. Second, the program OrthoFinder (version 2.3.1, Emms and Kelly 2015) was used to identify ortholog groups at the hierarchical level of the last common ancestor of the
254 investigated species. OrthoFinder generated a) a list of ortholog groups that contained one gene from each of the tested species only (single-copy genes) and b) rooted phylogenetic trees
256 of genes that belong to ortholog groups with paralogous gene copies. The software DIAMOND (Buchfink et al. 2014) was used to align the amino-acid sequences of a given ortholog group.

258 Third, we queried the WaspAtlas database (http://cyverse.warwick.ac.uk:3000/, Davies and Tauber 2015) for ortholog groups (Flicek et al. 2014).

260 We were unable to identify unambiguous orthologs of some of the *D. melanogaster* CHC biosynthesis candidate genes in the *N. vitripennis* genome with the above-outlined methods 262 despite partially high sequence similarities. Therefore, BLAST+-inferred hits in the *N. vitripennis* proteome with amino-acid sequence similarity thresholds of >-30-% identity, <-10^{-} 264 ^{10} e-value, and >-85-% query coverage to *D. melanogaster* were additionally considered, although these may not necessarily constitute direct orthologs. The identification of 266 orthologous candidate proteins of two protein families posed particular problems: fatty acid elongases (ELOs) and fatty-acyl-CoA reductases (FARs). Candidate ELOs and FARs in the *N.* 268 *vitripennis* proteome with the highest amino-acid sequence similarity (lowest e-value and highest bit score) to the *D. melanogaster* ELOs and FARs with demonstrated impact on CHC 270 variation (Tab. S2) were used as queries to search for homologs in the *Nasonia* proteome using blastp of the BLAST+ software suite. Hits above a similarity threshold of e-value < 10^{-20} were 272 considered to belong to the same protein/gene family, and domain structure of the respective hits was taken into account to ensure that the derived gene sequences encode functional 274 proteins. This was achieved with hmmscan of the HMMER software (http://hmmer.org/, version 3.1b2, February 2015) and use of the Pfam-A database (version 32, El-Gebali et al. 276 2019).

For visualization of the QTL regions and candidate gene positions on the *N. vitripennis* linkage 278 map Nvit 2.1, the R package LinkageMapView was used (Ouellette et al. 2017).

280

282 ***CHC profile discriminant analysis***

Discriminant analysis (DA) was performed with the R package "MASS" (Venables and Ripley
284 2002) to test whether CHC profiles statistically differ between the six investigated groups of male wasps (i.e., those of the parental strains and those of the F_2 hybrids). To standardize the
286 peak area values for DA, the normalization method of the function "decostand" of the community ecology R package "vegan" was used based on the following formula (Dixon 2003):

288

$$T_{x,y} = P_{x,y} / \sqrt{\Sigma P_y^2}$$

290

$T_{x,y}$ is the transformed peak area x of individual y, $P_{x,y}$ is the absolute peak area x of individual
292 y and ΣP_y^2 to the squared sums of all absolute peak areas of individual y. This widely applied method for normalizing phenotypic data was chosen to make the peak areas comparable
294 between our groups and to highlight the relative peak area differences. To visualize the data by plotting the first three discriminant functions simultaneously, the R package
296 "scatterplot3d" was used (Ligges and Maechler 2003). Wilk's λ was calculated to measure the quality of the DA.

298

300 ***Gene orthology and screens for signatures of positive selection***

To test for signatures of positive selection in our candidate genes, we extended our ortholog
302 inference analysis to the fourth *Nasonia* species, *N. oneida*, as well as two closely related parasitoid wasp genera, *Trichomalopsis sarcophagae* and *Muscidifurax raptorellus*. The
304 nucleotide sequences of the candidate genes from the *N. vitripennis* genome were extracted with SAMtools (Li et al. 2009) and aligned with the software BLAT (v35, Kent 2002; Bhagwat

306 et al. 2012) to the nucleotide sequences of the latest genome assemblies of *N. giraulti*, *N. longicornis*, *N. oneida*, and *T. sarcophagae* kindly provided by Xiaozhu Wang (unpublished
308 data). Aligned nucleotide sequences that shared at least 50 identical nucleotide positions and were aligned along at least 80 % query nucleotide sequence length were considered as
310 homologous. To ensure these were also orthologous to the candidate genes of the *N. vitripennis* genome, their amino-acid sequences were reciprocally searched against the
312 *N. vitripennis* proteome with BLAST+ (version 2.8.1, Camacho et al. 2009). Only genes that fulfilled the best reciprocal hit criterion were considered as orthologous. To annotate genes
314 in the draft genomes of *N. giraulti*, *N. oneida*, *N. longicornis*, and *T. sarcophagae*, we used the program GeMoMa (version 1.6, Keilwagen et al. 2016). The reliability of the annotation by the
316 program can be improved by providing it the coding sequence of an orthologous gene if a phylogenetically conserved gene structure can be assumed. We applied this approach for
318 structurally annotating candidate genes in the *N. giraulti*, *N. oneida*, *N. longicornis*, and *T. sarcophagae* genomes, using nucleotide sequences of the orthologous *N. vitripennis* genes
320 as references. Since BLAT was not sensitive enough to identify orthologs in *M. raptorellus* (comparatively distantly related to *N. vitripennis*, see Fig. 2), we *de novo*-inferred gene models
322 in a draft genome of this species, kindly provided by Eva Jongepier (unpublished data), with the software GeMoMa. To obtain the respective coding nucleotide sequences from the
324 GeMoMa-predicted annotations, the program BEDtools (Version 2.28.0, Quinlan and Hall 2010) with either the nucleotide sequences of the gene regions or the whole genome
326 assembly.

Candidate genes were screened for signatures of positive selection by analyzing the ratio of
328 nonsynonymous to synonymous substitutions (ω = dN/dS). The nucleotide sequences of each candidate gene ortholog group containing the orthologs of all six wasp species were aligned

330 with the codon-sensitive alignment tool PRANK (version 170427, Löytynoja and Goldman 2008; Löytynoja 2014). The alignment process was supported with a phylogenetic tree (Fig. 2) 332 of the study species whose evolutionary distances are based on the D2 expansion region of their 28S rDNA sequences (Campbell et al. 2000) and mitochondrial DNA sequences (Werren 334 et al. 2010). Poorly aligned regions were trimmed with Gblocks (version 0.91b, Castresana 2000). Each sequence alignment was screened for signatures of positive selection on at least 336 one site on at least one branch of the given phylogeny with the BUSTED (**B**ranch-**S**ite **U**nrestricted **S**tatistical **T**est for **E**pisodic **D**iversification) algorithm (Murrell et al. 2015), 338 implemented in HyPhy (version 2.5, http://hyphy.org). Resulting p values were Benjamini-Hochberg-corrected for multiple testing (Benjamini and Hochberg 1995).

340

342

RESULTS

344

QTL explaining CHC variation in Nasonia *F2 hybrid males*

346 We detected a total of 91 QTL explaining CHC variation in recombinant F_2 hybrid males. Of those, 80 were found in the recombinant F2 hybrid males obtained from crossing 348 *N. longicornis* (♂) and *N. vitripennis* (♀) and eleven were found in the recombinant F2 hybrid males obtained from crossing *N. longicornis* (♂) and *N. giraulti* (♀) (Fig. 3 and 4, Tab. 1). The 350 detected QTL explain variation in 43 (of a total of 53 analyzed) CHCs. The QTL of seven CHCs (two mono-methyl-branched, three di-methyl-branched, and two tetra-methyl-branched 352 alkanes) were found in hybrids from both crosses, but corresponding QTL are located on

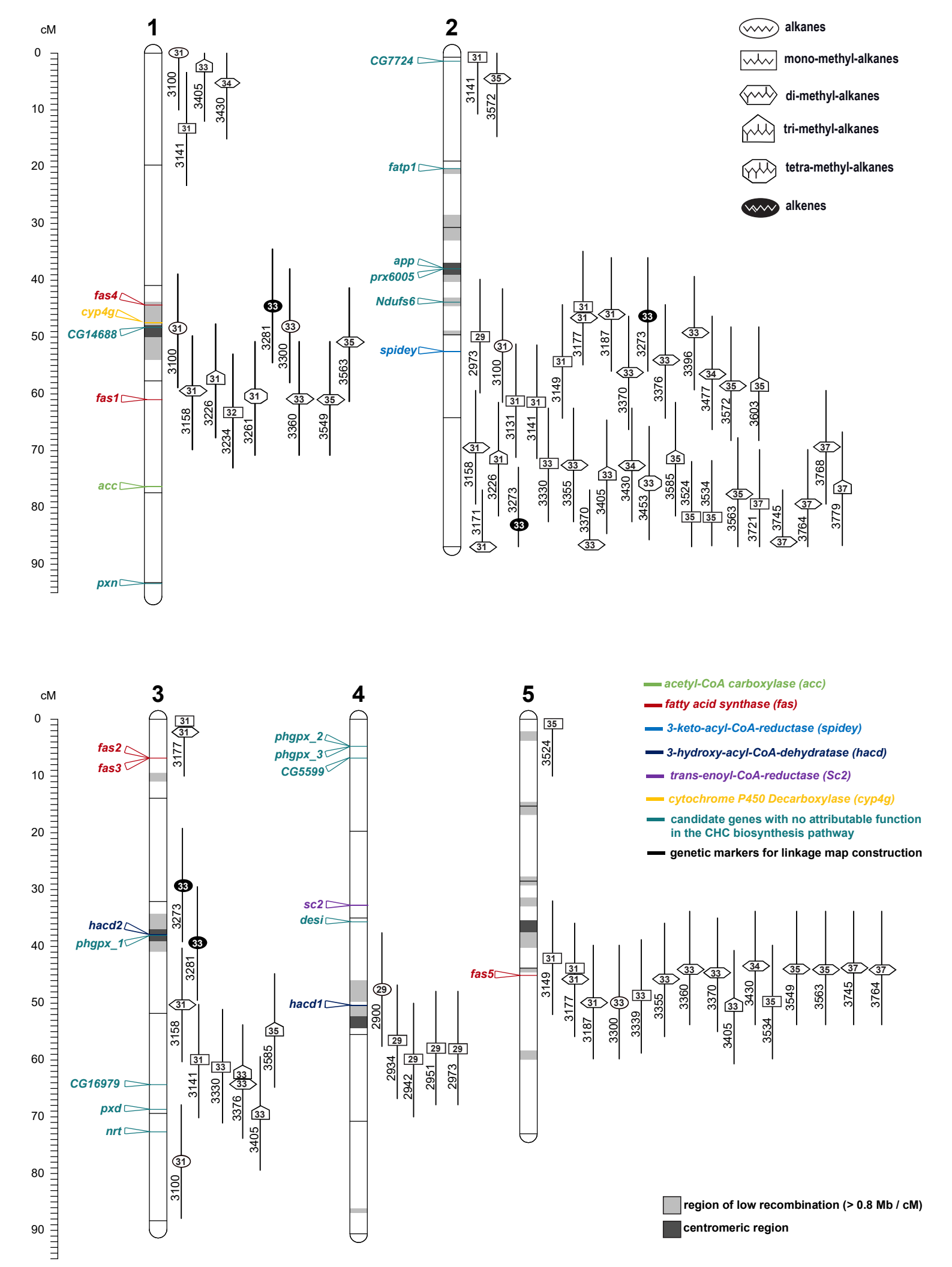
cM
1
2
3
4
5
alkanes
mono-methyl-alkanes
di-methyl-alkanes
tri-methyl-alkanes
tetra-methyl-alkanes
alkenes
fas4
cyp4g
CG14688
fas1
acc
pxn
CG7724
fatp1
app
prx6005
Ndufs6
spidey
fas2
fas3
hacd2
phgpx_1
CG16979
pxd
nrt
phgpx_2
phgpx_3
CG5599
sc2
desi
hacd1
fas5
acetyl-CoA carboxylase (acc)
fatty acid synthase (fas)
3-keto-acyl-CoA-reductase (spidey)
3-hydroxy-acyl-CoA-dehydratase (hacd)
trans-enoyl-CoA-reductase (Sc2)
cytochrome P450 Decarboxylase (cyp4g)
candidate genes with no attributable function in the CHC biosynthesis pathway
genetic markers for linkage map construction
region of low recombination (> 0.8 Mb / cM)
centromeric region

Figure 3: Linkage map based on F2 hybrid males from a cross between *Nasonia longicornis* (♂) and *Nasonia vitripennis* (♀) depicting positions of 80 quantitative trait loci (QTL) for 42 individual cuticular hydrocarbons (CHCs). Additionally, the positions of CHC biosynthesis-related candidate genes are shown. CHCs are indicated by symbols corresponding to their respective compound class (alkanes, alkenes, mono-, di-, tri-, and tetra-methyl-alkanes), their retention indices (RI) and numbers inside the symbols reflecting their carbon chain length. The five Nasonia chromosomes are labeled 1–5. Black horizontal lines on the chromosomes depict the respective positions of the 29 genotyped molecular markers used for determining the QTL positions. Orthologs to genes with a clear impact on CHC production derived from *Drosophila melanogaster* are color coded according to their putative position in the CHC biosynthesis pathway where possible (see Fig. 1). Regions of low recombination (> 0.8 Mb / cM) and centromeric regions are derived from Desjardins et al. (2013) and marked in light and dark grey, respectively.

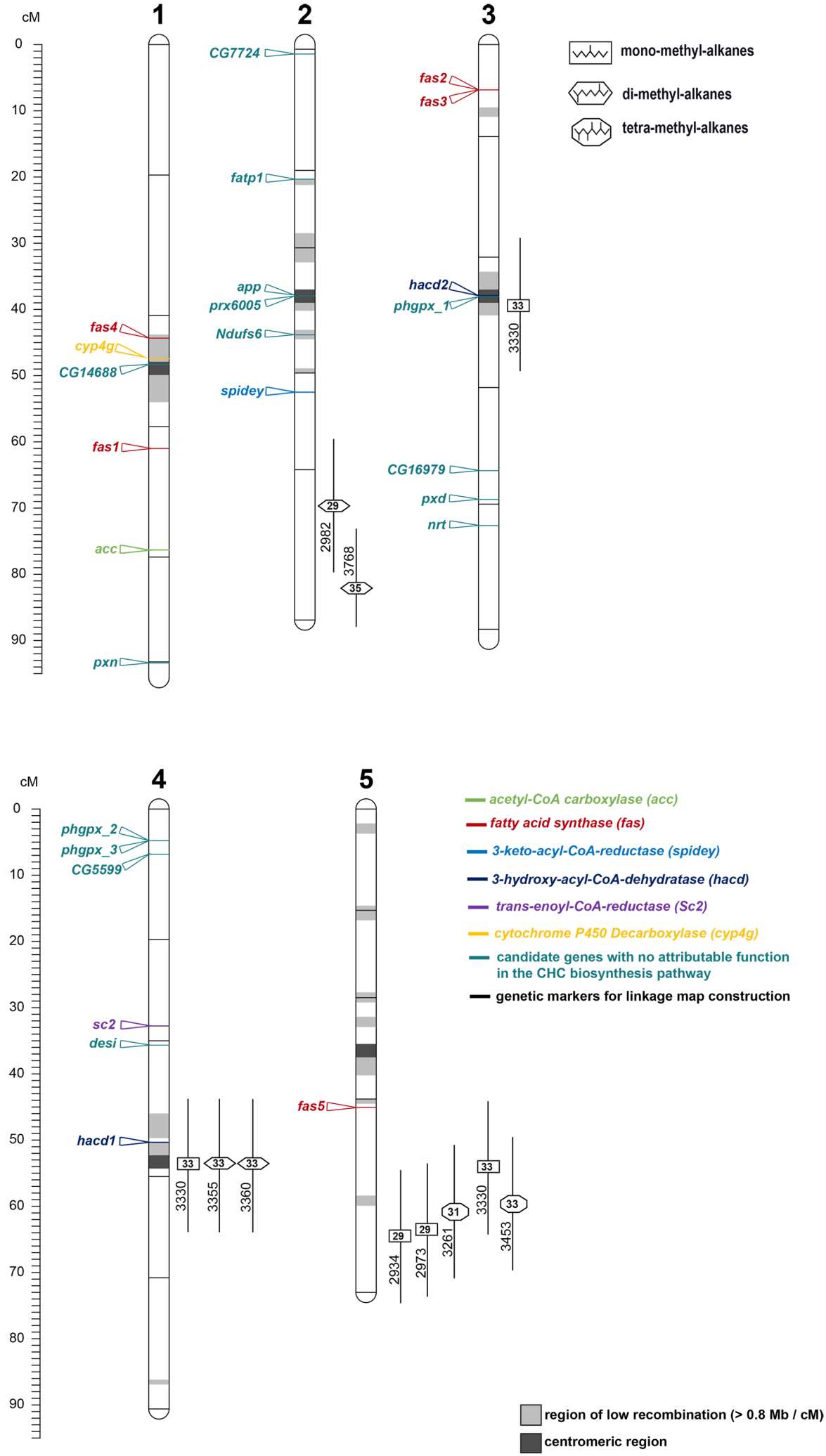
cM
1
2
3
4
5
0
10
20
30
40
50
60
70
80
90
fas4
cyp4g
CG14688
fas1
acc
pxn
CG7724
fatp1
app
prx6005
Ndufs6
spidey
29
2982
3768
35
fas2
fas3
hacd2
phgpx_1
33
3330
CG16979
pxd
nrt
mono-methyl-alkanes
di-methyl-alkanes
tetra-methyl-alkanes
phgpx_2
phgpx_3
CG5599
sc2
desi
hacd1
33
3330
33
3355
33
3360
fas5
29
2934
29
2973
31
3261
33
3330
33
3453
acetyl-CoA carboxylase (acc)
fatty acid synthase (fas)
3-keto-acyl-CoA-reductase (spidey)
3-hydroxy-acyl-CoA-dehydratase (hacd)
trans-enoyl-CoA-reductase (Sc2)
cytochrome P450 Decarboxylase (cyp4g)
candidate genes with no attributable function in the CHC biosynthesis pathway
genetic markers for linkage map construction
region of low recombination (> 0.8 Mb / cM)
centromeric region

Figure 4: Linkage map based on F2 hybrid males from a cross between *Nasonia longicornis* (♂) and *Nasonia giraulti* (♀) depicting positions of 11 quantitative trait loci (QTL) for 9 individual cuticular hydrocarbons (CHCs). Additionally, the positions of CHC biosynthesis-related candidate genes are shown. QTL could only be detected for the variation of methyl-branched alkanes (mono-, di-, and tetra-methyl-alkanes). Labels of chromosomes, candidate genes, molecular markers and regions of low recombination are indicated as in Fig. 3

different chromosomes or on different chromosomal regions (Tab. 1). The QTL of one single
354 CHC compound, 15,17-DiMeC29 (RI: 2982), was detected exclusively in hybrids that originated
from crossing *N. longicornis* and *N. giraulti* (Fig. 4 and Tab. 1). Sixty-nine QTL, explaining
356 variation of 35 CHCs, were exclusively detected in hybrids that originated from crossing
N. longicornis and *N. vitripennis* (Fig. 3 and Tab. 1). The QTL detected in F2 hybrid males of the
358 latter cross are spread across all five chromosomes and encompass all six compound classes
identified in *Nasonia* CHC profiles (i.e., alkanes, alkenes, mono-, di-, tri-, and tetra-methyl-
360 branched alkanes). Several QTL explaining quantitative variation of structurally similar CHCs
(mostly methyl-branched alkanes) clustered together, with two clusters each on
362 chromosomes 1 and 2, and one cluster each on chromosomes 3, 4, and 5 (Fig. 3 and Table 1).
Eleven QTL were found in recombinant F2 hybrid males obtained from crossing *N. longicornis*
364 and *N. giraulti* (Fig. 2). All of them explain variation of methyl-branched alkanes (mono-, di-,
and tetra-methy-alkanes). These QTL are located on chromosomes 2, 4, and 5, with clusters
366 of them detected on chromosomes 4 and 5 (Fig. 4). Although these QTL clusters are in spatial
proximity to the ones detected on the same two chromosomes in the F2 male hybrids
368 obtained from crossing *N. longicornis* and *N. vitripennis*, they explain variation of different
CHCs (compare Fig. 3 and 4).

370

372 ***CHC biosynthesis-related candidate genes***

Of the candidate genes from *Drosophila melanogaster* with a demonstrated impact on CHC
374 variation through targeted knockdown studies (Table S2), we found 15 to be represented by
exactly one ortholog in the *Nasonia* reference genome, irrespective of the applied ortholog
376 inference method (i.e., reciprocal BLAST search, OrthoFinder, and WaspAtlas) (Table 2).

Ortholog inference of the remaining 15 *D. melanogaster* candidate genes generated
378 inconsistent results when comparing the three ortholog inference methods. For instance, reciprocal BLAST search (RBS) of three fatty acid synthase (FAS) genes with a demonstrated
380 impact on *Drosophila* CHC biosynthesis identified nine putative *fas* orthologs in the *Nasonia* reference genome, whereas OrthoFinder found five. However, upon closer inspection, several
382 of the putative *fas* orthologs identified by RBS are actually comprised of very short nucleotide sequences that unlikely encode fully functional proteins. These potential pseudogenes were
384 excluded as CHC biosynthesis candidate genes in *Nasonia*. Furthermore, of all three ortholog inference methods, WaspAtlas was the only tool that led to the identification of an *N.*
386 *vitripennis* ortholog to the *D. melanogaster* gene *cyp4g1*. However, due to a complimentary identification of an orthologous member of the *Cyp4g* gene family in the *N. vitripennis* genome
388 with the highest nucleotide sequence similarity to *cyp4g1* by Qiu et al (2012), this ortholog was still included in the list of potential CHC biosynthesis candidate genes.
390 Two further gene families emerged in our analysis with high nucleotide sequence similarity to *Drosophila* CHC biosynthesis candidate genes despite no evidence for direct orthology:
392 *elongases* (*elos*) and *fatty-acyl-CoA reductases* (*fars*). Of the four *elos* and the two *fars* in the *D. melanogaster* genome with a clear impact on CHC profile composition, we were able to
394 identify twelve and 22 homologous genes with high nucleotide sequence similarity in the *Nasonia* reference genome (Tab. 3 and Fig. 5). Mapping of those genes on the *Nasonia* linkage
396 map revealed a region on chromosome 1 around the centromere where eight *elos* and thirteen *fars* cluster together in close spatial vicinity (Fig. 5). Two additional candidate genes,
398 orthologous to the *Drosophila* CHC biosynthesis genes *cyp4g1* and *CG14688*, also map in this chromosomal region.
400

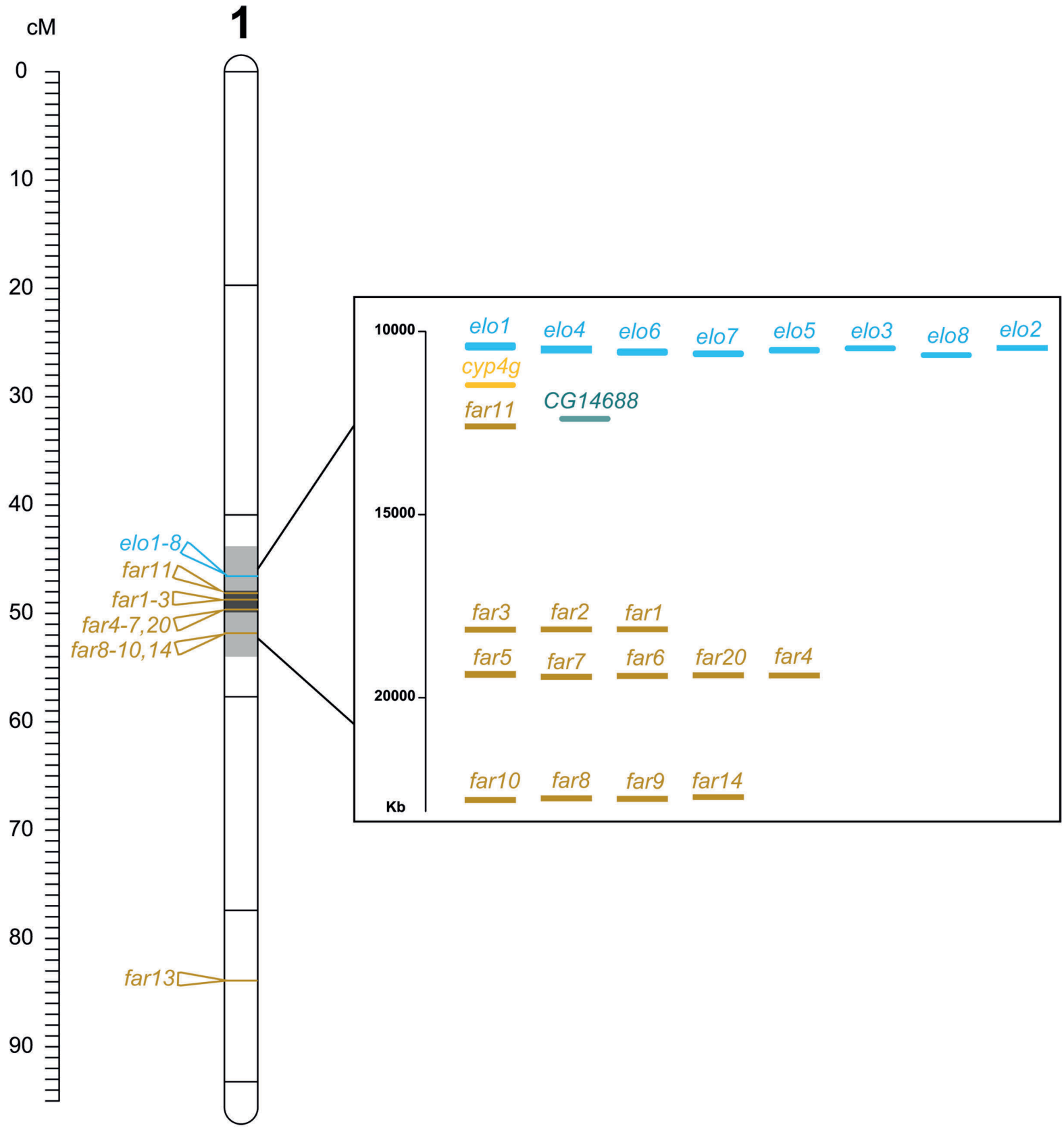
cM
1
0
10
20
30
40
50
60
70
80
90
elo1-8
far11
far1-3
far4-7,20
far8-10,14
far13
10000
15000
20000
Kb
elo1
elo4
elo6
elo7
elo5
elo3
elo8
elo2
cyp4g
far11
CG14688
far3
far2
far1
far5
far7
far6
far20
far4
far10
far8
far9
far14

Figure 5: *Nasonia* linkage map of chromosome 1 with an enlarged region densely packed with CHC biosynthesis related candidate genes. Eight elongase (elo) and thirteen fatty-acyl-reductase (far) cluster together in close spatial vicinity in addition to two additional genes genes cyp4g1 and CG14688 (see Tab. 2 and 3) in this chromosomal region. Note that this region maps around the centromere and is characterized by a particularly low recombination rate (> 0.8 Mb / cM). Labels of chromosomes and molecular markers are indicated as in Fig. 3 and 4.

402 ***Correlation in the genomic localization of candidate genes and QTL***

QTL from the second cluster on chromosome 1 explaining variation of CHCs in recombinant F_2
404 male hybrids of *N. longicornis* and *N. vitripennis* co-localize with CHC biosynthesis-related candidate genes in close spatial vicinity to the centromere (Fig. 3). Intriguingly, this also
406 encompasses the conspicuous chromosomal region with a high density of *elo* and *far* genes (Fig. 5). Of the two detected QTL clusters on chromosome 2 from the cross mentioned above,
408 the first one co-localizes with an ortholog of the candidate gene *spidey* that codes for a component of the elongase enzyme complex (Wicker-Thomas et al. 2015; Holze et al. 2021,
410 see Fig. 1 and 3). Some QTL from the second cluster co-localize with several *far* genes (compare Fig. 3 and Fig. 5). On chromosome 3, the detected QTL cluster maps closely to the
412 position of a gene orthologous to *CG16979*, a candidate gene with an impact on CHC variation in *D. melanogaster* despite an unclear role in the CHC biosynthesis pathway (Dembeck et al.
414 2015; Holze et al. 2021). On chromosome 4, several QTL map closely to a *hacd1*-ortholog, another gene that codes for an elongase enzyme complex component (Wicker-Thomas et al.
416 2015; Holze et al. 2021). On chromosome 5, a particularly large cluster, consisting almost entirely of QTL explaining variation in methyl-branched alkanes, co-localizes with a *fas* gene
418 ortholog (*fas5*). Another QTL cluster identified in male F_2 hybrids from the cross between *N. longicornis* and *N. giraulti* on chromosome 4 maps close to the *hacd1* ortholog, in close
420 proximity to the above-mentioned QTL cluster from the cross between *N. vitripennis* and *N. longicornis* on the same chromosome (compare Fig. 3 and 4).

422

424

CHC divergence between F_0 parental and F_2 hybrid males

426 To assess whether CHC profiles statistically differ between males of the parental strains and the recombinant F_2 hybrid males, we conducted a discriminant analysis (DA) on the CHC profile
428 data of 256 male wasps belonging to five groups: *N. giraulti* (NG), *N. longicornis* (NL), *N. vitripennis* (NV), *N. longicornis* x *N. vitripennis* F2 hybrids (LV), and *N. longicornis* x *N. giraulti*
430 F2 hybrids (LG) (Fig. 6). All groups differed in their CHC profiles significantly from each other (Wilk's $\lambda < 0.001$, $\chi^2 = 23.22$, $p < 0.001$), and the recombinant F_2 males consistently clustered
432 in between their respective F_0 parental males.

434

Orthology and signatures of positive selection for CHC candidate genes in other
436 ***parasitoid wasps***

We screened all 26 CHC biosynthesis-related candidate genes identified in the *N. vitripennis*
438 reference genome (see Tab. 2) for orthologs in the respective genomes of the other three *Nasonia* species and in the genomes of two additional pteromalid wasp species: *Muscidifurax*
440 *raptorellus* and *Trichomalopsis sarcophagae* (Fig. 2). In the genome of the latter, we were unable to identify orthologs of seven of the 26 *N. vitripennis* CHC biosynthesis candidate genes
442 (*CG14688, fas2, fas4, fas5, hacd2, nrt*). Of these seven, two (*fas5* and *hacd2*) remained unidentified in *T. sarcophagae* and three remained unidentified in *N. giraulti* (*fatp1, nrt,*
444 *hacd2*). Interestingly, *hacd2* was the only *N. vitripennis*-exclusive CHC biosynthesis-related candidate gene with no orthologs in any of the other investigated parasitoid wasp species
446 (Tab. 4). Orthologs of the remaining 19 CHC biosynthesis-related candidate genes were found in all investigated species and were screened for signatures of positive selection based on the

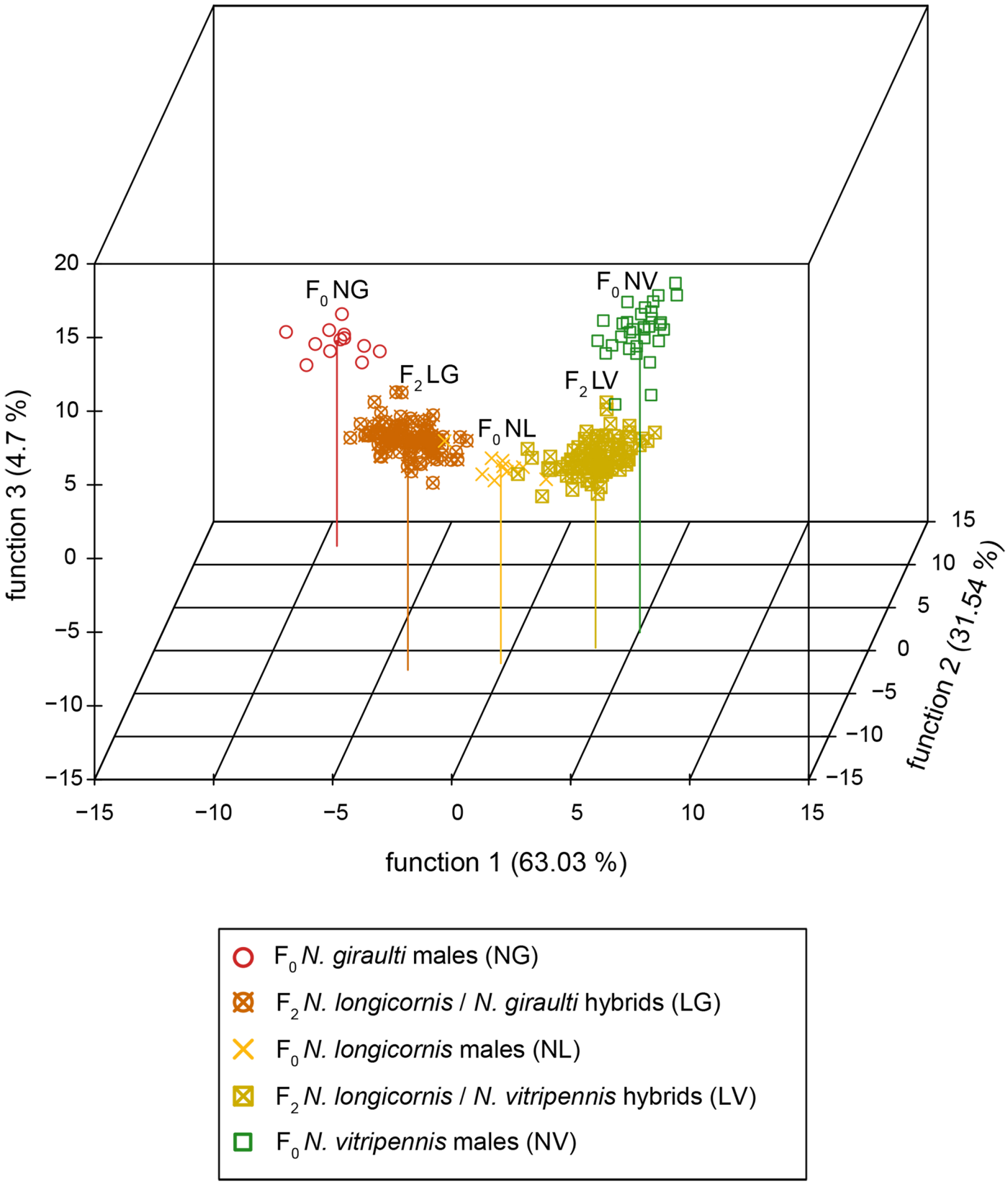
F0 NG
F2 LG
F0 NL
F2 LV
F0 NV
function 3 (4.7 %)
function 1 (63.03 %)
function 2 (31.54 %)
20
15
10
5
0
−5
−10
−15
F0 N. giraulti males (NG)
F2 N. longicornis / N. giraulti hybrids (LG)
F0 N. longicornis males (NL)
F2 N. longicornis / N. vitripennis hybrids (LV)
F0 N. vitripennis males (NV)

Figure 6: Plot of the first three discriminant functions showing the divergence between CHC profiles of F0 parental and F2 hybrid males between the three investigated *Nasonia* species in three dimensions. F0 NG: *N. giraulti* males (n = 12), F0 NL: *N. longicornis* males (n = 11), F0 NV: *N. vitripennis* males (n = 32), F2 LV: *N. longicornis* / *N. vitripennis* hybrid males (n = 101), F2 LG: *N. longicornis* / *N. giraulti* hybrid males (n = 100), total n = 256. All groups were significantly differentiated from each other (Wilk's $\lambda < 0.001$, $\chi 2 = 23.22$, $p < 0.001$), the variation each function explains is indicated in percentages.

ratios of non-synonymous (dN) to synonymous (dS) nucleotide substitutions. After Benjamini-Hochberg correction for multiple testing, the signature for positive selection in one gene, *fatty acid synthase 1* (*fas1*), was considered statistically significant.

DISCUSSION

Genomic regions harboring candidate genes and QTL explaining CHC variation

By analyzing recombinant F2 hybrid males from crosses between *N. longicornis* and *N. vitripennis* (LV) and between *N. longicornis* x *N. giraulti* (LG), we identified several QTL clusters mapping to genomic regions that harbor CHC biosynthesis candidate genes. Four of these regions additionally coincide with the location of QTL explaining CHC variation in recombinant F2 hybrid males of *N. vitripennis* and *N. giraulti* (VG, Niehuis et al. 2011, Fig. S1 and Tab. S3). The coinciding QTL of the latter study additionally explain variation in six CHCs (i.e., 7-MeC30, 5-MeC30, 9-C31ene, 3,7-DiMeC31, 3,25-DiMeC31, 3-MeC32, 15-; 17-; 19-MeC37) that we found no QTL of in the present investigation. Interestingly, QTL explaining variation of identical compounds in different hybrid crosses map only in very few instances to the same genomic regions (e.g., QTL for 11,15-; 11,23- 11,25-DiMeC33 on Chr. 5 in LV and in VG crosses, QTL for 7,19-; 7,21-; 7,23-DiMeC37 on Chr. 2 in LV and in LG crosses). Conversely, in the majority of cases, QTL explaining variation of the same CHC compounds map to different chromosomal regions, or to different chromosomes entirely (Fig. 3, 4 and S1). This illustrates

the genetic complexity and potential pleiotropy of the governance of most investigated CHC compounds. However, specifically comparing our cross between *N. vitripennis* and *N. longicornis* with the previously analyzed cross between *N. vitripennis* and *N. giraulti*, we identified overlapping QTL clusters governing the variation of structurally similar CHC compounds. The three most obvious of these clusters map close to the centromeric regions of chromosomes 1, 2, and 5 (compare Fig. 3 and S1, Tab. S3). The respective CHC compounds are all methyl-branched alkanes, partially sharing methyl-branching patterns and chain lengths (Tab. 1). Intriguingly, all of the above-mentioned QTL clusters co-localize with orthologs of candidate genes implicated with CHC biosynthesis. Particularly, the shared cluster on chromosome 1 mapped to the identified chromosomal region with the highest density of orthologs of CHC biosynthesis candidate genes (compare Fig. 3 and Fig. 5). As the recombination rate around the centromeres is very low, the close proximity of these gene orthologs is concordant with a higher likelihood of passing the whole unaltered gene cluster to the next generation, rendering gene shuffling in this region rare (Niehuis et al. 2010). The other two clusters co-localize with the positions of the gene orthologs *spidey* and *fas5* on chromosome 2 and 5, respectively. *Spidey* encodes a 3-hydroxy-acyl-CoA-dehydratase, constituting a part of the elongase enzyme complex (Wicker-Thomas et al. 2015), whereas *fas5* encodes a fatty acid synthase, instrumental for the formation of long-chain fatty acids as precursors for the elongation process (Juarez et al. 1992; Gu et al. 1997, see Fig. 2). These findings render these three distinct chromosomal regions particularly promising targets for further investigating the impact of the candidate genes mapped in these regions on CHC biosynthesis.

CHC compound class variation and gene orthology between Nasonia and Drosophila

The vast majority of QTL that we detected in the hybrids from both crosses explained variation of methyl-branched alkanes (79 of 91 in total). This compound class dominates in CHC profiles of *Nasonia* wasp males and accounts for 87 % of the total amount of CHCs when averaged across all three investigated species (Fig. 1 and Tab. S4). This result is in accordance with those of most studies on CHC profiles in Hymenoptera, consistently reporting methyl-branched alkanes as the dominant CHC class (e.g. Martin and Drijfhout 2009; Kather and Martin 2015; Sprenger and Menzel 2020). In species of other insect orders, however, methyl-branched alkanes seem to be less dominant, which is particularly true for *Drosophila melanogaster*: analyzing CHC profiles in lines of the *D. melanogaster* Reference Panel (DGRP), Dembeck et al. (2015) detected only mono-methyl-branched alkanes in *D. melanogaster* CHC profiles, and these account for 16.2 % and 24.1 % of the total CHCs in CHC profiles of males and females, respectively (Tab. S5). This might be one of the reasons why our understanding of the genetics governing the variation in methyl-branched alkanes is comparatively limited.

The split between Hymenoptera and other holometabolous insects, including Diptera, is estimated to have occurred around 327 Mya (Misof et al. 2014). Thus, it is not surprising that were we unable to reliably identify orthologs of several genes with an impact on CHC profile composition in *Drosophila* in the genome of *Nasonia* wasps. For instance, the potentially best studied CHC biosynthesis-related genes in *Drosophila* are desaturases, catalyzing the introduction of double bounds in hydrocarbon chains (e.g. Wicker-Thomas et al. 1997; Coyne et al. 1999; Dallerac et al. 2000, Fig. 1). Although several *desaturase* genes have been identified in the *Nasonia* genome (Niehuis et al. 2011), we were not able to clearly confirm their orthology to *Drosophila* desaturases in the present study. However, the proportions of

unsaturated CHCs vastly differ between the CHC profiles of *Drosophila* flies and those of
520 *Nasonia* wasps (on average 59.2 % vs. 1.4 % in males, Tab. S4 and S5). Thus, since the proportion of unsaturated compounds is considerably less in *Nasonia* CHC profiles (compared
522 to CHC profiles of *Drosophila* flies), the main genes governing their biosynthesis and variation might have deviated considerably after more than 300 Mya of estimated evolutionary
524 divergence between *Nasonia* and *Drosophila* (Misof et al. 2014).

Similar in their unresolved orthology to *Drosophila*, we found numerous *elo* and *far* genes
526 across the *Nasonia* genome (Tab. 3). In *D. melanogaster*, 19 *elo* genes have been identified in total (Chung and Carroll 2015; Zuo et al. 2018), but only four have been clearly associated with
528 CHC biosynthesis and variation (Chertemps et al. 2007; Dembeck et al. 2015). Concerning *far* genes, from 17 identified in the *D. melanogster* genome (Finet et al. 2019), two have a
530 demonstrated impact on CHC biosynthesis (Dembeck et al. 2015). This already indicates the difficulty in associating members in these two large and diverse gene families with CHC
532 biosynthesis and variation, as they can also be involved in vastly different processes, including tracheal gas exchange, reproductive tissue growth, and larval development (Jaspers et al.
534 2014; Cinnamon et al. 2016). However, there are two further properties of *Nasonia* CHC profiles that greatly differentiate them from their *Drosophila* counterparts: first, in *Drosophila*,
536 CHC chain lengths apparently do not exceed beyond C31 (Dembeck et al. 2015), whereas in *Nasonia*, CHC compounds with chain lengths of up to C37 have been identified, and additional
538 CHCs with chain length of up to C52 have recently been described (Bien et al. 2019). Since elongases contribute to the elongation of long chain fatty acids (see Fig. 1), the longer and
540 more diverse chain lengths in the *Nasonia* profiles could mean that a larger and more diverse set of *elongase* genes is functionally involved in CHC biosynthesis in *Nasonia* wasps. Second,
542 *Nasonia* wasp CHC profiles appear to be more complex, generally containing more compound

classes, including di-, tri-, tetra-methyl-branched alkanes, in contrast to CHC profiles of
544 *D. melanogaster* (compare Tab. S4 and S5). It has been demonstrated that the *far* gene family shows particularly quick evolutionary turnover rates, which allows, in turn, for rapid
546 diversification of CHC profiles between *Nasonia* species (Finet et al. 2019). This also argues for the involvement of a larger set of *far* genes in CHC biosynthesis in *Nasonia* than in *Drosophila*
548 due to the more diverse and complex nature of *Nasonia* CHC profiles. Hence, the region on chromosome 1 containing eight *elo* and eleven *far* genes while harboring QTL mostly
550 explaining methyl-branched alkane variation in two of our analyzed hybrid crosses constitutes a promising target-rich region for future functional genetic studies. *Nasonia* might prove to be
552 a particularly well-suited model system for investigating the genetic underpinnings of methyl-branched alkane biosynthesis and diversification, with our current study providing a
554 comprehensive foundation.

556

Phylogenetic divergence in relation to CHC differentiation and selection
558 ***signatures***

It has been estimated that *N. vitripennis* has diverged around 1 Mya from the lineage from
560 which the remaining *Nasonia* species evolved (Campbell et al. 1993). The divergences of the remaining *Nasonia* lineages occurred approximately between 500 000 and 400 000 years ago
562 (Werren et al. 2010, Fig. 2). We discovered the smallest number of QTL governing CHC variation between the more recently diverged species pair studied by us: *N. longicornis* and
564 *N. giraulti*. This finding matches well with the considerably shorter evolutionary time frame to accumulate genetic differences compared to the longer divergence time between each of the
566 two species and *N. vitripennis*. Therefore, it is all the more surprising that phenotypically, the

overall CHC profile divergence sufficiently discriminates all species with F_2 hybrid males
568 clustering as intermediate phenotypes between the respective F_0 males (Fig. 6). As
N. longicornis and *N. giraulti* males also differ considerable in their CHC composition from
570 each other, this obvious phenotypic divergence is not reflected in the comparably low
genomic divergence between these two species hinted at by our QTL comparison. A potential
572 explanation for this might be other mechanisms, *e.g.*, phenotypic plasticity (Otte et al. 2018)
or epigenetic regulation (Morandin et al. 2019), that factor into CHC profile divergence
574 between these phylogenetically less distant species.

Concerning CHC biosynthesis candidate gene orthology, we extended our analysis beyond the
576 genus *Nasonia* to two more distantly related parasitoid wasp species in the family
Pteromalidae, *Trichomalopsis sarcophagae and Muscidifurax raptorellus* (Fig. 2).
578 Unsurprisingly, in the genome of the latter, most distantly related species, we detected the
fewest orthologs in comparison to the *N. vitripennis* reference genome (19 out of 26, Tab. 4).
580 Furthermore, taking into account all shared 19 candidate genes in our six investigated wasp
genomes, we detected signatures of positive selection in only one of them, *fas1*. The gene
582 *fas1*, an ortholog to *FASN1* and *FASN2* in *Drosophila*, plays an important role in the early steps
of CHC biosynthesis and impacts the abundance of methyl-branched alkanes (Chung et al.
584 2014; Wicker-Thomas et al. 2015). Additionally, *fas1* maps in close proximity to the
conspicuous cluster of CHC biosynthesis gene orthologs (Fig. 5) that also co-localizes with QTL
586 explaining variation of mostly methyl-branched alkanes between *N. vitripennis* and *N. giraulti*
and between *N. longicornis* and *N. giraulti* (compare Fig. 3 and S1). This renders *fas1* a
588 particularly promising candidate gene for further functional genetic studies investigating its
impact on CHC profiles not only in *Nasonia*, but also in other Hymenoptera. Generally,
590 establishing a functional link between the identified candidate genes and CHC variation will

be instrumental in further elucidating the genomic and genetic architecture underlying species-specific CHC diversification in parasitoid wasps in particular and other hymenopteran species in general.

Conclusions and Outlook

With our study on the genetic and genomic architecture of CHC variation between parasitoid wasps of the genus *Nasonia*, we uncovered several genomic regions where QTL explaining variation in similar CHC compounds and CHC biosynthesis-related candidate genes clustered. Especially one region on chromosome 1 close to the centromere is conspicuous in this regard, as it harbors many CHC biosynthesis-related candidate genes and QTL explaining variation of methyl-branched alkanes, the prevalent CHC compound class in the Hymenoptera in general and in *Nasonia* in particular. This underlines the considerable potential of *Nasonia* as a model system to further investigate the little-know genetic and genomic architecture of methyl-branched alkane biosynthesis and variation.

ACKNOWLEDGEMENTS

Thanks to Andrea K. Judson, Nadine Brehm and Sylvia Geeritsma for their help in acquiring the chemical and QTL data, Joshua D. Gibson for his valuable input in conceptualizing the manuscript in its current form, and Valerio Vitali for helpful suggestions on the first draft of the manuscript.

FUNDING

This research was partially funded by the Deutsche Forschungsgemeinschaft (DFG, German Research Foundation) – 427879779; 403980864.

CONFLICT OF INTEREST

The authors declare no conflict of interest.

DATA AVAILABILITY STATEMENT

All data underlying the presented study will be made available at the dryad data repository

REFERENCES

Benjamini, Y. and Y. Hochberg. 1995. Controlling the false discovery rate: A practical and powerful approach to multiple testing. J. Roy. Stat. Soc. Ser. B. (Stat. Method.). 57(1): 289-300.

Bhagwat, M., L. Young and R. R. Robison. 2012. Using BLAT to find sequence similarity in closely related genomes. Current protocols in bioinformatics. 37: 10.18.11-10.18.24.

638 Bien, T., J. Gadau, A. Schnapp, J. Y. Yew, C. Sievert and K. Dreisewerd. 2019. Detection of very long-chain hydrocarbons by laser mass spectrometry reveals novel species-, sex-
640 , and age-dependent differences in the cuticular profiles of three *Nasonia* species. Anal. Bioanal. Chem. 411(13): 2981-2993.

642

Blomquist, G. J. and A. G. Bagnères. 2010. Insect hydrocarbons: Biology, biochemistry, and
644 chemical ecology. Cambridge University Press, Cambridge, Massachusetts.

646 Blomquist, G. J. and M. D. Ginzel. 2021. Chemical Ecology, Biochemistry, and Molecular Biology of Insect Hydrocarbons. Annu. Rev. Entomol. 66(1): 45-60.

648

Blomquist, G. J., D. R. Nelson and M. Derenobales. 1987. Chemistry, biochemistry, and
650 physiology of insect cuticular lipids. Arch. Insect Biochem. Physiol. 6(4): 227-265.

652 Bordenstein, S. R., F. P. O'Hara and J. H. Werren. 2001. *Wolbachia*-induced incompatibility precedes other hybrid incompatibilities in *Nasonia*. Nature. 409(6821): 707.

654

Bordenstein, S. R. and J. H. Werren. 1998. Effects of A and B *Wolbachia* and host genotype
656 on interspecies cytoplasmic incompatibility in *Nasonia*. Genetics. 148(4): 1833-1844.

658 Breeuwer, J. A. J. and J. H. Werren. 1990. Microorganisms associated with chromosome destruction and reproductive isolation between two insect species. Nature. 346: 558-
660 560.

662 Breeuwer, J. A. J. and J. H. Werren. 1995. Hybrid breakdown between two haplodiploid species: The role of nuclear and cytoplasmic genes. Evolution. 49: 705-717.

664

Broman, K. W. and S. Sen. 2009. A Guide to QTL Mapping with R/QTL. Springer, Dordrecht.

666

Broman, K. W., H. Wu, S. Sen and G. A. Churchill. 2003. R/qtl: QTL mapping in experimental

668 crosses. Bioinformatics. 19(7): 889–890.

670 Buchfink, B., C. Xie and D. H. Huson. 2014. Fast and sensitive protein alignment using DIAMOND. Nat. Methods. 12: 59–60.

672

Buellesbach, J., J. Gadau, L. W. Beukeboom, F. Echinger, R. Raychoudhury, J. H. Werren and

674 T. Schmitt. 2013. Cuticular hydrocarbon divergence in the jewel wasp *Nasonia*: Evolutionary shifts in chemical communication channels? J. Evol. Biol. 26(11): 2467-

676 2478.

678 Camacho, C., G. Coulouris, V. Avagyan, N. Ma, J. Papadopoulos, K. Bealer and T. L. Madden. 2009. BLAST+: architecture and applications. BMC Bioinformatics. 10: 421.

680

Campbell, B., J. Heraty, J.-Y. Rasplus, K. Chan, J. Steffen-Campbell and C. Babock. 2000.

682 Molecular Systematics of the Chalcidoidea using 28S-D2 rDNA, in Hymenoptera: Evolution, Biodiversity and Biological Control edited by A. D. Austin and M. Dowton.

684 CISRO Publishing, Collingwood.

686 Campbell, B. C., J. D. Steffen-Campbell and J. H. Werren. 1993. Phylogeny of the Nasonia species complex (Hymenoptera: Pteromalidae) inferred from an internal transcribed
688 spacer (ITS2) and 28S rDNA sequences. Insect Biochem. Mol. Biol. 2(4): 225-237.

690 Carlson, D. A., U. R. Bernier and B. D. Sutton. 1998. Elution patterns from capillary GC for methyl-branched alkanes. J. Chem. Ecol. 24(11): 1845-1865.

692

Carlson, D. A., M. S. Mayer, D. L. Silhacek, J. D. James, M. Beroza and B. A. Bierl. 1971. Sex
694 attractant pheromone of the housefly: Isolation, identification, and synthesis. Science. 174: 76-77.

696

Castresana, J. 2000. Selection of conserved blocks from multiple alignments for their use in
698 phylogenetic analysis. Mol. Biol. Evol. 17(4): 540–552.

700 Chertemps, T., L. Duportets, C. Labeur, R. Ueda, K. Takahashi, K. Saigo and C. Wicker-Thomas. 2007. A female-biased expressed elongase involved in long-chain
702 hydrocarbon biosynthesis and courtship behavior in *Drosophila melanogaster*. Proc. Natl. Acad. Sci. USA. 104(11): 4273-4278.

704

Chung, H. and S. B. Carroll. 2015. Wax, sex and the origin of species: Dual roles of insect
706 cuticular hydrocarbons in adaptation and mating. Bioessays. 37(7): 822-830.

708 Chung, H., D. W. Loehlin, H. D. Dufour, K. Vaccarro, J. G. Millar and S. B. Carroll. 2014. A single gene affects both ecological divergence and mate choice in *Drosophila*.
710 Science. 343(6175): 1148-1151.

712 Cinnamon, E., R. Makki, A. Sawala, L. P. Wickenberg, G. J. Blomquist, C. Tittiger, Z. e. Paroush and A. P. Gould. 2016. *Drosophila* Spidey/Kar regulates oenocyte growth via PI3-
714 kinase signaling. PLoS Genet. 12(8): e1006154.

716 Coyne, J. A. and R. Oyama. 1995. Localization of pheromonal sexual dimorphism in *Drosophila melanogaster* and its effect on sexual isolation. Proc. Natl. Acad. Sci. USA.
718 92(21): 9505-9509.

720 Coyne, J. A., C. Wicker-Thomas and J. M. Jallon. 1999. A gene responsible for a cuticular hydrocarbon polymorphism in *Drosophila melanogaster*. Genet. Res. 73(3): 189-203.
722

Dallerac, R., C. Labeur, J. M. Jallon, D. C. Knippie, W. L. Roelofs and C. Wicker-Thomas. 2000.
724 A Δ9 desaturase gene with a different substrate specificity is responsible for the cuticular diene hydrocarbon polymorphism in *Drosophila melanogaster*. Proc. Natl.
726 Acad. Sci. USA. 97(17): 9449-9454.

728 Darling, D. C. and J. H. Werren. 1990. Biosystematics of *Nasonia* (Hymenoptera: Pteromalidae): Two new species reared from birds' nests in North America. Annals of
730 the Entomological Society of America. 83: 352-370.

732 Davies, N. J. and E. Tauber. 2015. WaspAtlas: a *Nasonia vitripennis* gene database and analysis platform. Database. bav103: 1-7.

734

Dembeck, L. M., K. Böröczky, W. Huang, C. Schal, R. R. H. Anholt, T. F. C. Mackay and D. J.
736 Kliebenstein. 2015. Genetic architecture of natural variation in cuticular hydrocarbon composition in *Drosophila melanogaster*. eLife. 4: e09861.

738

Desjardins, C. A., J. Gadau, J. A. Lopez, O. Niehuis, A. R. Avery, D. W. Loehlin, S. Richards, J. K.
740 Colbourne and J. H. Werren. 2013. Fine-scale mapping of the Nasonia genome to chromosomes using a high-density genotyping microarray. G3: Genes, Genomes,
742 Genetics. 3(2): 205-215.

744 Dixon, P. 2003. VEGAN, a package of R functions for community ecology. J. Veg. Sci. 14(6): 927-930.

746

El-Gebali, S., J. Mistry, A. Bateman, S. R. Eddy, A. Luciani, S. C. Potter, M. Qureshi, L. J.
748 Richardson, G. A. Salazar, A. Smart et al. 2019. The Pfam protein families database in 2019. Nucleic Acids Research. 47(D1): D427-D432.

750

Emms, D. M. and S. Kelly. 2015. OrthoFinder: solving fundamental biases in whole genome
752 comparisons dramatically improves orthogroup inference accuracy. Genome Biology. 16(1): 157.

754

Erickson, D. L., C. B. Fenster, H. K. Stenøien and D. Price. 2004. Quantitative trait locus analyses and the study of evolutionary process. Mol. Ecol. 13(9): 2505-2522.

Ferveur, J.-F. 2005. Cuticular hydrocarbons: Their evolution and roles in *Drosophila* pheromonal communication. Behav. Genet. 35(3): 279-295.

Finet, C., K. Slavik, J. Pu, S. B. Carroll and H. Chung. 2019. Birth-and-death evolution of the fatty acyl-CoA reductase (FAR) gene family and diversification of cuticular hydrocarbon synthesis in *Drosophila*. Genome Biol. Evol. 11(6): 1541–1551.

Flicek, P., M. R. Amode, D. Barrell, K. Beal, K. Billis, S. Brent, D. Carvalho-Silva, P. Clapham, G. Coates, S. Fitzgerald et al. 2014. Ensembl 2014. Nucleic acids research. 42: D749-D755.

Gadau, J., O. Niehuis, A. Peire, J. H. Werren, E. Baudry and L. W. Beukeboom. 2008. The jewel wasp, *Nasonia* spp., p. 27-41 in Genome Mapping and Genomics in Animals edited by W. Hunter and C. Kole. Springer, Berlin Heidelberg.

Gadau, J., R. E. Page and J. H. Werren. 2002. The genetic basis of the interspecific differences in wing size in *Nasonia* (Hymenoptera: Pteromalidae): Major quantitative trait loci and epistasis. Genetics. 161: 673-684.

Gramates, L. S., S. J. Marygold, G. d. Santos, J.-M. Urbano, G. Antonazzo, B. B. Matthews, A. J. Rey, C. J. Tabone, M. A. Crosby, D. B. Emmert et al. 2017. FlyBase at 25: looking to the future. Nucleic Acids Research. 45(D1): D663-D671.

Gu, P., W. H. Welch, L. Guo, K. M. Schegg and G. J. Blomquist. 1997. Characterization of a novel microsomal fatty acid synthetase (FAS) compared to a cytosolic FAS in the housefly, *Musca domestica*. Comp. Biochem. Physiol. B-Biochem. Mol. Biol. 118(2): 447–456.

Haley, C. S. and S. A. Knott. 1992. A simple regression method for mapping quantitative trait loci in line crosses using flanking markers. Heredity. 69: 315-324.

Hartl, D. 2020. Population genetics of complex traits, p. 225-261 in A primer of population genetics and genomics edited by. 4 ed. Oxford University Press, Oxford.

Hollander, M. and D. A. Wolfe. 1999. Nonparametric Statistical Methods. Wiley, New York.

Holze, H., L. Schrader and J. Buellesbach. 2021. Advances in deciphering the genetic basis of insect cuticular hydrocarbon biosynthesis and variation. Heredity. 126(2): 219-234.

Howard, R. W. and C. Blomquist. 1982. Chemical ecology and biochemistry of insect hydrocarbons. Annu. Rev. Entomol. 27: 149-172.

800 Howard, R. W. and G. J. Blomquist. 2005. Ecological, behavioral, and biochemical aspects of insect hydrocarbons. Annu. Rev. Entomol. 50: 371-393.

802

Hoy, M. A. 2005. Behavior of Insects: Genetic Analysis by Crossing and Selection, p. 278-281

804 in Encyclopedia of Entomology edited by J. L. Capinera. Springer, Dordrecht.

806 Hunter, W. and C. Kole. 2008. Genome Mapping and Genomics in Arthropods. Springer, Berlin Heidelberg.

808

Jallon, J. M. 1984. A few chemical words exchanged by *Drosophila* during courtship and

810 mating. Behav. Genet. 14(5): 441-478.

812 Jaspers, M. H. J., R. Pflanz, D. Riedel, S. Kawelke, I. Feussner and R. Schuh. 2014. The fatty acyl-CoA reductase Waterproof mediates airway clearance in *Drosophila*. Dev. Biol.

814 385(1): 23–31.

816 Juarez, P., J. Chase and G. J. Blomquist. 1992. A microsomal fatty acid synthetase from the integument of *Blattella germanica* synthesizes methyl-branched fatty acids,

818 precursors to hydrocarbon and contact sex pheromone. Arch. Biochem. Biophys. 293(2): 333–341.

820

Kather, R. and S. J. Martin. 2015. Evolution of cuticular hydrocarbons in the Hymenoptera: A

822 meta-analysis. J. Chem. Ecol. 41(10): 871-883.

824 Keilwagen, J., M. Wenk, J. L. Erickson, M. H. Schattat, J. Grau and F. Hartung. 2016. Using intron position conservation for homology-based gene prediction. Nucleic Acids
826 Research. 44(9): e89.

828 Kent, W. J. 2002. BLAT--the BLAST-like alignment tool. Genome Res. 12(4): 656–664.

830 Lander, E. S. and D. Botstein. 1989. Mapping mendelian factors underlying quantitative traits using RFLP linkage maps. Genetics. 121(1): 185-199.

832

Li, H., B. Handsaker, A. Wysoker, T. Fennell, J. Ruan, N. Homer, G. Marth, G. Abecasis and R.
834 Durbin. 2009. The Sequence Alignment/Map format and SAMtools. Bioinformatics. 25(16): 2078–2079.

836

Ligges, U. and M. Maechler. 2003. Scatterplot3d – An R package for visualizing multivariate
838 data. J. Stat. Soft. 8(11): 1-20.

840 Lockey, K. H. 1980. Insect cuticular hydrocarbons. Comp. Biochem. Physiol. B: Biochem. Mol. Biol. 65(3): 457-462.

842

Lockey, K. H. 1988. Lipids of the insect cuticle: origin, composition and function. Comp.
844 Biochem. Physiol. B: Biochem. Mol. Biol. 89(4): 595-645.

846 Löytynoja, A. 2014. Phylogeny-aware alignment with PRANK, p. 155-170 in Multiple Sequence Alignment Methods edited by D. J. Russell. Humana Press, Totowa, NJ.

Löytynoja, A. and N. Goldman. 2008. Phylogeny-aware gap placement prevents errors in sequence alignment and evolutionary analysis. Science. 320(5883): 1632–1635.

Mackay, T. F. C. 2001a. The Genetic Architecture of Quantitative Traits. Annu. Rev. Genet. 35(1): 303-339.

Mackay, T. F. C. 2001b. Quantitative trait loci in *Drosophila*. Nature Reviews Genetics. 2(1): 11-20.

Mackay, T. F. C., E. A. Stone and J. F. Ayroles. 2009. The genetics of quantitative traits: challenges and prospects. Nature Reviews Genetics. 10(8): 565-577.

Manichaikul, A., J. Y. Moon, S. Sen, B. S. Yandell and K. W. Broman. 2009. A model selection approach for the identification of quantitative trait loci in experimental crosses, allowing epistasis. Genetics. 181(3): 1077-1086.

Martin, S. and F. Drijfhout. 2009. A review of ant cuticular hydrocarbons. J. Chem. Ecol. 35(10): 1151-1161.

Misof, B., S. Liu, K. Meusemann, R. S. Peters, A. Donath, C. Mayer, P. B. Frandsen, J. Ware, T. Flouri, R. G. Beutel et al. 2014. Phylogenomics resolves the timing and pattern of insect evolution. Science. 346(6210): 763-767.

872 Morandin, C., V. P. Brendel, L. Sundstrom, H. Helantera and A. S. Mikheyev. 2019. Changes in gene DNA methylation and expression networks accompany caste specialization and
874 age-related physiological changes in a social insect. Mol. Ecol. 28(8): 1975-1993.

876 Murrell, B., S. Weaver, M. D. Smith, J. O. Wertheim, S. Murrell, A. Aylward, K. Eren, T. Pollner, D. P. Martin, D. M. Smith et al. 2015. Gene-wide identification of episodic
878 selection. Mol. Biol. Evol. 32(5): 1365–1371.

880 Nelson, D. R. and G. J. Blomquist. 1995. Insect waxes, p. 1-90 in Waxes: Chemistry, molecular biology and functions edited by R. J. Hamilton. Oily Press, Dundee, Scotland.

882

Niehuis, O., J. Buellesbach, A. K. Judson, T. Schmitt and J. Gadau. 2011. Genetics of cuticular
884 hydrocarbon differences between males of the parasitoid wasps *Nasonia giraulti* and *Nasonia vitripennis*. Heredity. 107(1): 61-70.

886

Niehuis, O., J. D. Gibson, M. S. Rosenberg, B. A. Pannebakker, T. Koevoets, A. K. Judson, C. A.
888 Desjardins, K. Kennedy, D. Duggan, L. W. Beukeboom et al. 2010. Recombination and its impact on the genome of the haplodiploid parasitoid wasp *Nasonia*. PloS one.
890 5(1): e8597.

892 Niehuis, O., A. K. Judson and G. J. 2008. Cytonuclear genic incompatibilities cause increased mortality in male F2 hybrids of *Nasonia giraulti* and *N. vitripennis*. Genetics. 178(1):
894 413-426.

896 Niehuis, O., A. K. Judson, J. H. Werren, W. B. Hunter, P. M. Dang, S. E. Dowd, B. Grillenberger, L. W. Beukeboom and J. Gadau. 2007. Species-diagnostic single-nucleotide
898 polymorphism and sequence-tagged site markers for the parasitic wasp genus *Nasonia* (Hymenoptera: Pteromalidae). J. Econ. Entomol. 100(4): 1033-1036.

900

Orr, A. H. 2001. The genetics of species differences. Trends Ecol. Evol. 16(7): 343-350.

902

Otte, T., M. Hilker and S. Geiselhardt. 2018. Phenotypic plasticity of cuticular hydrocarbon
904 profiles in insects. J. Chem. Ecol. 44(3): 235-247.

906 Ouellette, L. A., R. W. Reid, S. G. Blanchard and C. R. Brouwer. 2017. LinkageMapView - Rendering High Resolution Linkage and QTL Maps. Bioinformatics. 34(2): 306-307.

908

Quinlan, A. R. and I. M. Hall. 2010. BEDTools: a flexible suite of utilities for comparing
910 genomic features. Bioinformatics. 26(6): 841–842.

912 R Development Core Team. 2018. R: A language and environment for statistical computing. Vienna, Austria, R foundation for statistical computing, http://www.R-project.org.

914

Ramsay, J. A. 1935. The evaporation of water from the cockroach. J. Exp. Biol. 12: 373-383.

916

Raychoudhury, R., C. A. Desjardins, J. Buellesbach, D. W. Loehlin, B. K. Grillenberger, L. W.
918 Beukeboom, T. Schmitt and J. H. Werren. 2010. Behavioral and genetic characteristics of a new species of *Nasonia*. Heredity. 104(3): 278-288.

920

Richardson, P. M., W. P. Holmes and G. B. Saul. 1987. The effect of tetracycline on
922 nonreciprocal cross incompatibility in *Mormoniella* [= *Nasonia*] vitripennis. J.
Invertebr. Pathol. 50(3): 176-183.

924

Slate, J. 2005. Quantitative trait locus mapping in natural populations: progress, caveats and
926 future directions. Mol. Ecol. 14(2): 363-379.

928 Sprenger, P. P. and F. Menzel. 2020. Cuticular hydrocarbons in ants (Hymenoptera:
Formicidae) and other insects: How and why they differ among individuals, colonies,
930 and species. Myrmecol. News. 30: 1-26.

932 Thomas, G. W. C., E. Dohmen, D. S. T. Hughes, S. C. Murali, M. Poelchau, K. Glastad, C. A.
Anstead, N. A. Ayoub, P. Batterham, M. Bellair et al. 2020. Gene content evolution in
934 the arthropods. Genome Biology. 21(1): 15.

936 Venables, W. N. and B. D. Ripley. 2002. Modern Applied Statistics with S. Springer, New York.

938 Werren, J. H. and D. W. Loehlin. 2009. The parasitoid wasp *Nasonia*: An emerging model
system with haploid male genetics. Cold Spring Harbor Protocols. 10.

940

Werren, J. H., S. Richards, C. A. Desjardins, O. Niehuis, J. Gadau, J. K. Colbourne and The
942 Nasonia Genome Working Group. 2010. Functional and evolutionary insights from
the genomes of three parasitoid *Nasonia* species. Science. 327(5963): 343-348.

944

Wicker-Thomas, C., D. Garrido, G. Bontonou, L. Napal, N. Mazuras, B. Denis, T. Rubin, J.-P. Parvy and J. Montagne. 2015. Flexible origin of hydrocarbon/pheromone precursors in *Drosophila melanogaster*. J. Lipid Res. 56(11): 2094-2101.

948

Wicker-Thomas, C., C. Henriet and R. Dallerac. 1997. Partial characterization of a fatty acid desaturase gene in *Drosophila melanogaster*. Insect Biochem. Mol. Biol. 27(11): 963-972.

952

Yang, J., J. Zhu and R. W. Williams. 2007. Mapping the genetic architecture of complex traits in experimental populations. Bioinformatics. 23(12): 1527-1536.

956 Zar. 1999. Biostatistical Analysis. Prentice-Hall, Englewood Cliffs, NJ.

958 Zuo, W., C. Li, Y. Luan, H. Zhang, X. Tong, M. Han, R. Gao, H. Hu, J. Song, F. Dai et al. 2018. Genome-wide identification and analysis of elongase of very long chain fatty acid genes in the silkworm, Bombyx mori. Genome. 61(3): 167–176.

TABLES

Tab. 1: Cuticular hydrocarbons (CHCs) identified in male *Nasonia* profiles and the corresponding QTL from crosses between *Nasonia longicornis* and *Nasonia vitripennis* as well as between *N. longicornis* and *N. girauli*. Indicated are the retention indices (RI), compound identifications or possible configurations in case of ambiguities, chromosome number (Chr.), QTL position (pos.), confidence interval for the respective QTL position, hybrid cross (LV or LG), ratios of the CHC compounds in parental males of the two species for each respective cross (percentages with standard deviations) and significance assessments comparing the two ratios (Benjamini Hochberg-corrected Mann-Whitney U tests).

RI	Compound identification / possible configurations	Chr.	QTL pos.	Confidence interval	Cross	NL ♂ ratio	NV ♂ ratio	NG ♂ ratio	Significance
2900	C29	4	47.6	37.6 - 57.6	LV	1.11 ± 0.44	0.98 ± 0.5		n.s.
2934	9-; 11-; 13-; 15-MeC29	4	56.8	46.8 - 66.8	LV	0.13 ± 0.21	0.08 ± 0.12		n.s.
		5	64.2	54.2 - 73	LG	0.13 ± 0.21		0.01 ± 0.05	0.02346
2942	7-MeC29	4	60	50 - 70	LV	0.78 ± 0.69	0.6 ± 0.31		n.s.
2951	5-MeC29	4	57.9	47.9 - 67.9	LV	0.36 ± 0.28	0.37 ± 0.16		n.s.
2973	3-MeC29	2	49.6	39.6 - 59.6	LV	0.15 ± 0.19	0.11 ± 0.06		n.s.
		4	57.9	47.9 - 67.9	LV				
		5	63.3	53.3 - 73	LG	0.15 ± 0.20		0.04 ± 0.07	n.s.
2982	15,17-DiMeC29	2	71.5	61.5 - 81.5	LG	0.09 ± 0.1		0.01 ± 0.03	n.s.
3100	C31	1	0	0 - 10	LV				
		1	48.9	38.9 - 58.9	LV				
		2	49.6	39.6 - 59.6	LV				
		3	77.9	67.9 - 87.9	LV				
3131	9-; 11-; 13-; 15-MeC31	2	61.2	51.2 - 71.2	LV	2.13 ± 0.89	1.99 ± 0.95		n.s.
3141	7-MeC31	1	13.3	3.3 - 23.3	LV	22.71 ± 4.49	11.43 ± 1.11		< 0.0001
		2	0.7	0 - 10.7	LV				
		2	54.3	44.3 - 64.3	LV				
		3	60.2	50.2 - 70.2	LV				
3149	5-MeC31	2	53.3	43.3 - 63.3	LV	7.17 ± 1.8	4.42 ± 0.65		0.00021
		5	42	32 - 52	LV				
3158	11,15-; 11,17-; 11,21-DiMeC31	2	69.4	59.4 - 79.4	LV	0.37 ± 0.12	0.46 ± 0.13		n.s.
		3	50.3	40.3 - 60.3	LV				
3171	7,11-DiMeC31	2	86.9	76.9 - 86.9	LV	4.84 ± 1.06	3.73 ± 0.35		0.01025

RI	Compound identification / possible configurations	Chr.	QTL pos.	Confidence interval	Cross	NL ♂ ratio	NV ♂ ratio	NG ♂ ratio	Significance
3177	3-MeC31 (+7,23-DiMeC31)	2	46	36 - 56	LV	3.42 ± 0.73	2.55 ± 0.55		n.s.
		5	49.8	39.8 - 59.8	LV				
3171	7,11-DiMeC31	2	86.9	76.9 - 86.9	LV	4.84 ± 1.06	3.73 ± 0.35		0.01025
3187	7,25-DiMeC31	2	44.9	34.9 - 54.9	LV	0.93 ± 0.2	0.61 ± 0.18		0.00655
		3	0	0 - 10	LV				
		5	45.9	35.9 - 55.9	LV				
3226	3,7,9-; 3,7,11-; 3,7,15-TriMeC31	1	57.7	47.7 - 67.7	LV	0.23 ± 0.09	0.16 ± 0.08		n.s.
		2	71.5	61.5 - 81.5	LV				
3234	6-MeC32	1	63	53 - 73	LV	0.47 ± 0.14	0.39 ± 0.05		n.s.
3261	3,7,11,15-TetraMeC31	1	60.8	50.8 - 70.8	LV	0.53 ± 0.35	0.33 ± 0.1		n.s.
		5	59.4	49.4 - 69.4	LG	0.53 ± 0.35		0.67 ± 0.18	n.s.
3273	9-C33ene	2	46	36 - 56	LV	0.31 ± 0.18	1.45 ± 0.27		< 0.0001
		2	82.9	72.9 - 86.9	LV				
		3	29.2	19.2 - 39.2	LV				
3281	7-C33ene	1	44.5	34.5 - 54.5	LV	0.17 ± 0.09	0.96 ± 0.2		< 0.0001
		3	39.5	29.5 - 49.5	LV				
3300	C33	1	48	38 - 58	LV	0.41 ± 0.2	0.99 ± 0.25		< 0.0001
		5	49.8	39.8 - 59.8	LV				
3330	9-; 11-; 13-; 15-MeC33	2	72.5	62.5 - 82.5	LV	4.98 ± 1.11	3.97 ± 1.26		n.s.
		3	61.1	51.1 - 71.1	LV				
		3	33.2	23.2 - 43.2	LG	4.98 ± 1.11		8.02 ± 0.92	0.00044
		4	55.5	45.5 - 65.5	LG				
		5	52.7	42.7 - 62.7	LG				
3339	7-MeC33	5	48.8	38.8 - 58.8	LV	3.07 ± 1.03	2.93 ± 0.68		n.s.
3355	11,15-; 11,23- 11,25-DiMeC33	2	72.5	62.5 - 82.5	LV	1.97 ± 0.87	0.66 ± 0.42		0.00059
		5	45.9	35.9 - 55.9	LV				
		4	55.5	45.5 - 65.5	LG				
3360	7,19-; 7,21-DiMeC33	1	59.8	49.8 - 69.8	LV	1.23 ± 0.36	1.68 ± 0.56		n.s.
		5	43.8	33.8 - 53.8	LV				
		4	55.5	45.5 - 65.5	LG	1.23 ± 0.36		1.24 ± 0.21	n.s.
3370	7,23-DiMeC33	2	56.3	46.3 - 66.3	LV	7.52 ± 1.63	7.71 ± 1.21		n.s.
		2	86.9	76.9 - 86.9	LV				
		5	45	35 - 55	LV				

RI	Compound identification / possible configurations	Chr.	QTL pos.	Confidence interval	Cross	NL ♂ ratio	NV ♂ ratio	NG ♂ ratio	Significance
3376	5,9-; 5,15-; 5,17-; 5,19-DiMeC33 (+11,15,23-TriMeC33)	2	54.3	44.3 - 64.3	LV	4.4 ± 1.22	5.85 ± 0.81		0.01573
		3	63.8	53.8 - 73.8	LV				
3396	3,9-; 3,15-; 3,17-DiMeC33	2	49.3	39.3 - 59.3	LV	0.78 ± 0.64	1.09 ± 0.23		n.s.
3405	5,9,11-; 5,9,15-TriMeC33	1	2	0 - 12	LV	0.04 ± 0.1	2.04 ± 0.68		< 0.0001
		2	74.6	64.6 - 84.6	LV				
		3	69.4	59.4 - 79.4	LV				
		5	50.7	40.7 - 60.7	LV				
3430	8,10-; 8,12-; 8,14-; 8,16-; 8,18-DiMeC34	1	5.1	0 - 15.1	LV	0.32 ± 0.14	0.92 ± 0.28		< 0.0001
		2	72.5	62.5 - 82.5	LV				
		5	43.8	33.8 - 53.8	LV				
3453	3,7,11,15-TetraMeC33	2	75.7	65.7 - 85.7	LV	0.89 ± 0.4	1.57 ± 0.35		0.00211
		5	58.4	48.4 - 68.4	LG	0.89 ± 0.4		4 ± 0.71	0.00011
3477	7,11-; 7,13-; 7,15-DiMeC34	2	56.3	46.3 - 66.3	LV	0.32 ± 0.19	0.41 ± 0.17		n.s.
3524	13-; 15-;17-MeC35	2	81.9	71.9 - 86.9	LV	1.65 ± 0.51	2.58 ± 0.44		0.00086
		5	0	0 - 10	LV				
3534	7-MeC35	2	81.9	71.9 - 86.9	LV	0.36 ± 0.1	0.77 ± 0.18		< 0.0001
		5	49.8	39.8 - 59.8	LV				
3549	11,17-; 11,19-; 11,21-; 11,23-; 13,17-; 13,19-; 13,21-; 13,17-; 13,19-; 13,21-;13,23-; 15,17-; 15,19-; 15,21-; 15,23-DiMeC35	1	60.8	50.8 - 70.8	LV	4.35 ± 0.85	3.08 ± 0.92		0.02070
		5	43.8	33.8 - 53.8	LV				
3563	7,19-; 7,21-; 7,23-DiMeC35	1	51.5	41.5 - 61.5	LV	2.33 ± 0.68	7.07 ± 1.47		< 0.0001
		2	77.7	67.7 - 86.9	LV				
		5	43.8	33.8 - 53.8	LV				
3572	5,15-; 5,17-; 5,19-; 5,21-; 5,23-DiMeC35	2	4.7	0 - 14.7	LV	2.84 ± 0.58	5.21 ± 0.73		< 0.0001
		2	58.2	48.2 - 68.2	LV				
3585	7,15,21-; 7,15,23-; 7,17,21-; 7,17,23-; 7,19,21-; 7,19,23-TriMeC35	2	71.5	61.5 - 81.5	LV	0.71 ± 0.34	1.04 ± 0.29		n.s.
		3	54.8	44.8 - 64.8	LV				
3603	5,9,13-; 5,9,15-; 5,9,17-; 5,9,19-; 5,9,21-; 5,11,13-; 5,11,15-; 5,11,17-; 5,11,19-; 5,11,21-TriMeC35	2	58.2	48.2 - 68.2	LV	0.42 ± 0.3	1.65 ± 0.4		< 0.0001
3721	13-; 15-; 17-; 19-MeC37	2	79.8	69.8 - 86.9	LV				
3745	11,19-; 11,21-; 11,23-; 11,25-; 13,19-; 13,21-; 13,23-; 15,19-; 15,21-; 15,23-DiMeC37	2	86.9	76.9 - 86.9	LV	1.48 ± 0.4	1.38 ± 0.46		n.s.
		5	43.8	33.8 - 53.8	LV				

RI	Compound identification / possible configurations	Chr.	QTL pos.	Confidence interval	Cross	NL ♂ ratio	NV ♂ ratio	NG ♂ ratio	Significance
3764	7,19-; 7,21-; 7,23-DiMeC37	1	60.8	50.8 - 70.8	LV	0.58 ± 0.27	2.39 ± 0.69		< 0.0001
		2	79.8	69.8 - 86.9	LV				
		5	43.8	33.8 - 53.8	LV				
		2	82.9	72.9 - 86.9	LG	0.58 ± 0.27		0.61 ± 0.16	n.s.
3768	5,15-; 5,17-; 5,23-DiMeC37	2	69.4	59.4 - 79.4	LV	1.03 ± 0.46	1.42 ± 0.35		n.s.
3779	7,15,17-; 7,15,19-; 7,15,23-; 7,19,17-; 7,19,19-; 7,19,23-TriMeC37	2	76.7	66.7 - 86.7	LV	0.12 ± 0.14	0.53 ± 0.36		0.00336

Tab. 2: Candidate genes for CHC biosynthesis and variation in *Nasonia vitripennis* based on orthology towards *Drosophila melanogaster,* where the impact on CHC profiles of the respective genes has been demonstrated through targeted knockdown studies. Indicated are the gene annotations, gene IDs obtained from RefSeq, descriptions of the gene function obtained from NCBI, linkage map positions (centimorgan (cM) on the respective chromosome (c)), recombination rates (Mb/cM), and evidence of orthology based on the three methods reciprocal blast search (RB), OrthoFinder (OF) and WaspAtlas (WA). Genes for which exactly one ortholog could be unambiguously identified are indicated with a '+', in other cases the method(s) that have generated the most consistently reliable results in each case are shown. Genes that can clearly be assigned to a particular step in CHC biosynthesis (compare to Fig. 1) are indicated in bold.

Gene annotation	RefSeq ID	Description NCBI	Linkage map position	Recomb. rate (Mb/cM)	Evidence of orthology
acc	**100123347**	acetyl-CoA carboxylase	c1: 75.9 cM	0.1114	+
app	100121627	palmitoyltransferase ZDHHC18	c2: 38 cM	13.3806	+
CG14688	100120956	phytanoyl-CoA dioxygenase	c1: 48.2 cM	20.6198	+
CG16979	100119053	ufm1-specific protease 2-like	c3: 64.2 cM	0.1015	+
CG5599	100122501	lipoamide acyltransferase component ofbranched-chain alpha-keto acid dehydrogenase complex, mitochondrial	c4: 6.6 cM	0.2967	+
CG7724	100120136	3 beta-hydroxysteroid dehydrogenase/Delta 5-->4-isomerase type 4	c2: 1.5 cM	0.5931	+
Ndufs6	100119717	NADH dehydrogenase (ubiquinone) Fe-Sprotein 6, 13kDa (NADH-coenzyme Q reductase)	c2: 43.8 cM	1.1653	+
cyp4g44	**100113546**	cytochrome P450 4G44	c1: 47.5 cM	2.6119	WA
desi	100115157	uncharacterized LOC100115157	c4: 35.8 cM	0.1267	+

Gene annotation	RefSeq ID	Description NCBI	Linkage map position	Recomb. rate (Mb/cM)	Evidence of orthology
fas1	**100121447**	fatty acid synthase	c1: 60.6 cM	0.3184	OF
fas2	**100122099**	fatty acid synthase-like	c3: 5.8 cM	0.556	OF
fas3	**100122083**	fatty acid synthase-like	c3: 5.8 cM	0.556	OF
fas4	**100119597**	fatty acid synthase-like	c1: 49.6 cM	20.78	OF
fas5	**100678737**	fatty acid synthase-like	c5: 45.3 cM	0.7646	OF
fatp1	**100120211**	long-chain fatty acid transport protein 4-like	c2: 20.4 cM	0.8866	RB
hacd1	**100119292**	very-long-chain (3R)-3-hydroxyacyl-CoAdehydratase hpo-8	c4: 50.4 cM	4.7456	+
hacd2	**100115280**	very-long-chain (3R)-3-hydroxyacyl-CoAdehydratase	c3: 38 cM	11.8001	+
nrt	100121165	neurotactin	c3: 73 cM	0.1823	+
phgpx_1	100677983	probable phospholipid hydroperoxideglutathione peroxidase	c3: 38 cM	11.8001	RB,WA
phgpx_2	100114603	probable phospholipid hydroperoxideglutathione peroxidase	c4: 4.4 cM	0.2203	RB,WA
phgpx_3	100123141	phospholipid hydroperoxide glutathioneperoxidase, mitochondrial-like	c4: 4.4 cM	0.2203	RB,WA
prx6005	100117827	peroxiredoxin-6	c2: 38 cM	13.3806	
pxd	100119885	peroxidase	c3: 68.6 cM	0.3195	OF,WA
pxn	100116337	peroxidasin	c1: 93.4 cM	0.1259	+
sc2	**107980451**	very-long-chain enoyl-CoA reductase	c4: 32.1 cM	0.1186	+
spidey	**100121752**	very-long-chain 3-oxoacyl-CoA reductase	c2: 52.6 cM	0.3081	+

Tab. 3: *Elongase* (*ELO*) and *fatty acyl-CoA reductase* (*FAR*) genes in the *Nasonia* reference genome homologous to *D. melanogaster ELOs* and *FARs* with a demonstrated impact on CHC profiles (*ELO*s: *CG18609*, *CG30008*, *CG9458*, *eloF*; *FAR*s: *CG10097*, *CG13091;* see Tab. S2). Indicated are the gene annotations, gene and respective protein IDs obtained from RefSeq, linkage map positions and recombination rates (Kb/cM).

Gene annotation	RefSeq Gene ID	RefSeq Protein IDs	Linkage map position	Recombination rate (Kb/cM)
elo1	100114999	XP_016845887.1 XP_008214189.1 XP_008214188.1 XP_001599838.1	c1: 46.7 cM	3.6211
elo2	100115035	XP_001599867.1	c1: 46.7 cM	3.6211
elo3	100115068	XP_003428006.1	c1: 46.7 cM	3.6211
elo4	100115103	XP_008214193.1 XP_001599914.2 XP_016845633.1	c1: 46.7 cM	3.6211
elo5	100115131	XP_001599942.1	c1: 46.7 cM	3.6211
elo6	100115199	XP_001599996.2 XP_016838633.1	c1: 46.7 cM	3.6211
elo7	100115239	XP_016838697.1 XP_001600017.1	c1: 46.7 cM	3.6211
elo8	100115278	XP_001600048.1	c1: 46.7 cM	3.6211
elo9	100116202	XP_008205141.1	c3: 42.3 cM	0.3664
elo10	100118619	XP_008205429.1	c4: 47.5 cM	0.8598
elo11	100120094	XP_001603768.2	c3: 38 cM	11.8001
elo12	100123468	XP_001607111.2	c5: 16.1 cM	1.8931
far1	100115559	XP_008213286.1	c1: 48.9 cM	22.4845
far2	100115595	XP_008213287.2	c1: 48.9 cM	22.4845
far3	100115642	XP_001600309.1	c1: 48.9 cM	22.4845
far4	100115711	XP_001601168.1	c1: 49.6 cM	20.7846
far5	100117683	XP_016845554.1 XP_016845553.1 XP_016845552.1 XP_016845551.1 XP_008205703.1	c1: 49.6 cM	20.7846
far6	100117801	XP_008205709.1 XP_001601942.1	c1: 49.6 cM	20.7846
far7	100117837	XP_001601970.1	c1: 49.6 cM	20.7846
far8	100118857	XP_008212002.1 XP_008212003.1	c1: 51.8 cM	2.0759
far9	100118893	XP_001602762.2	c1: 51.8 cM	2.0759
far10	100119003	XP_001602857.1	c1: 51.8 cM	2.0759
far11	100121470	XP_016841641.1 XP_016841638.1 XP_016841633.1	c1: 48.2 cM	20.6198

Gene annotation	RefSeq Gene ID	RefSeq Protein IDs	Linkage map position	Recombination rate (Kb/cM)
far12	100121766	XP_008217013.1	c5: 19 cM	0.2367
far13	100123790	XP_001607507.3	c1: 83.2 cM	0.5953
far14	103317077	XP_016836513.1	c1: 51.8 cM	2.0759
far15	100117218	XP_001601521.2	c2: 85.4 cM	0.1672
far16	100117254	XP_001601550.4	c2: 85.4 cM	0.1672
far17	100117104	XP_001601438.1	c2: 85.4 cM	0.1672
far18	100117143	XP_001601466.1	c2: 85.4 cM	0.1672
far19	100117180	XP_001601494.1	c2: 85.4 cM	0.1672
far20	100117760	XP_016836801.1 XP_016836800.1	c1: 49.6 cM	20.7846
far21	100679544	XP_003426741.2	c2: 86.9 cM	NA
far22	100117878	XP_016843183.1	c3: 38 cM	0.1544

Tab. 4.: Orthologous candidate genes for CHC biosynthesis and variation in the *Nasonia* species complex as well as two representative species from the two most closely related genera, *Trichomalopsis sarcophagae* and *Muscidifurax raptorellus* (see Fig. 2 for phylogeny). Genes with identified orthologs in all analyzed species were subjected to a d_N/d_S ratio analysis to screen for signatures of positive selection with BUSTED. Indicated are gene annotations and gene IDs obtained from the *N. vitripennis* reference genome, species IDs where no orthologous gene could be unambiguously identified, and Benjamini-Hochberg-corrected p-values from the BUSTED analysis for signatures of positive selection. Species abbreviations: MR: *M. raptorellus*, TS: *T. sarcophagae*, NG: *N. giraulti*, NO: *N. oneida*, NL: *N. longicornis*.

Nv gene annotation	Nv gene ID	Positive selection (BUSTED, *p*-value)	Species with no identifiable ortholog
acc	100123347	0.081	
app	100121627	0.5	
CG14688	100120956		MR
CG16979	100119053	0.283	
CG5599	100122501	0.073	
CG7724	100120136	0.5	
Ndufs6	100119717	0.5	
cyp4g44	100113546	0.5	
desi	100115157	0.5	
fas1	**100121447**	**0.013**	
fas2	100122099		MR

Nv gene annotation	Nv gene ID	Positive selection (BUSTED, *p*-value)	Species with no identifiable ortholog
fas3	100122083	0.5	
fas4	100119597		MR
fas5	100678737		TS, MR
fatp1	100120211		NG, MR
hacd1	100119292	0.5	
hacd2	100115280		NO, NG, NL, TS, MR
nrt	100121165		NG, MR
phgpx_1	100677983	0.5	
phgpx_2	100114603	0.5	
phgpx_3	100123141	0.5	
prx6005	100117827	0.363	
pxd	100119885	0.5	
pxn	100116337	0.071	
sc2	107980451	0.363	
spidey	100121752	0.5	