

Supplementary tables for: “Candidate genes involved in cuticular hydrocarbon differentiation between cryptic, parabiotic ant species”

Philipp P. Sprenger, Juliane Hartke, Thomas Schmitt, Florian Menzel, Barbara Feldmeyer

Table S1: List of colonies sampled. The table includes the colony ID, species, sampling location and GPS coordinates of the parabiotic colonies collected initially and their experimental usage (RNA = used for RNAseq and transcriptome assembly; CHC = used for chemical analysis).

Colony ID	Species	Location	GPS	Usage
063-PAR	<i>Cr. levior</i> A	Paracou	N 05°15.574 W 52°55.815	RNA + CHC
064-PAR	<i>Cr. levior</i> A	Paracou	N 05°15.552 W 52°55.852	RNA + CHC
075-PAR	<i>Cr. levior</i> A	Paracou	N 05°15.639 W 52°56.097	RNA + CHC
077-PAR	<i>Cr. levior</i> A	Paracou	N 05°15.698 W 52°55.913	CHC only; library preparation failed
231-PAT	<i>Cr. levior</i> A	Camp Patawa	N 04°31.945 W 52°07.391	RNA only
240-PAT	<i>Cr. levior</i> A	Camp Patawa	N 04°33.241 W 52°09.040	RNA + CHC
242-PAT	<i>Cr. levior</i> A	Camp Patawa	N 04°33.281 W 52°09.065	RNA + CHC
059-PAR	<i>Cr. levior</i> B	Paracou	N 05°15.344 W 52°55.617	RNA + CHC
069-PAR	<i>Cr. levior</i> B	Paracou	N 05°15.503 W 52°55.982	RNA + CHC
074-PAR	<i>Cr. levior</i> B	Paracou	N 05°15.578 W 52°56.133	RNA + CHC
076-PAR	<i>Cr. levior</i> B	Paracou	N 05°15.698 W 52°55.913	RNA + CHC
218-PAT	<i>Cr. levior</i> B	Camp Patawa	N 04°32.703 W 52°07.991	RNA only
219-PAT	<i>Cr. levior</i> B	Camp Patawa	N 04°32.702 W 52°07.946	RNA only
222-PAT	<i>Cr. levior</i> B	Camp Patawa	N 04°32.687 W 52°07.872	RNA + CHC
230-PAT	<i>Cr. levior</i> B	Camp Patawa	N 04°31.572 W 52°07.107	RNA + CHC
232-PAT	<i>Cr. levior</i> B	Camp Patawa	N 04°32.269 W 52°07.923	RNA + CHC
045-PAR	<i>Ca. femoratus</i> PAT	Paracou	N 05°16.671 W 52°55.110	group dead upon arrival
051-PAR	<i>Ca. femoratus</i> PAT	Paracou	N 05°16.097 W 52°55.611	RNA only
059-PAR	<i>Ca. femoratus</i> PAT	Paracou	N 05°15.344 W 52°55.617	RNA only
077-PAR	<i>Ca. femoratus</i> PAT	Paracou	N 05°15.698 W 52°55.913	RNA + CHC
218-PAT	<i>Ca. femoratus</i> PAT	Patawa	N 04°32.703 W 52°07.991	RNA + CHC
232-PAT	<i>Ca. femoratus</i> PAT	Camp Patawa	N 04°32.269 W 52°07.923	RNA only

240-PAT	<i>Ca. femoratus</i> PAT	Camp Patawa	N 04°33.241 W 52°09.040	RNA only
242-PAT	<i>Ca. femoratus</i> PAT	Camp Patawa	N 04°33.281 W 52°09.065	RNA only
053-PAR	<i>Ca. femoratus</i> PS	Paracou	N 05°16.163 W 52°55.447	RNA + CHC
055-PAR	<i>Ca. femoratus</i> PS	Paracou	N 05°15.702 W 52°55.443	RNA + CHC
063-PAR	<i>Ca. femoratus</i> PS	Paracou	N 05°15.574 W 52°55.815	RNA + CHC
064-PAR	<i>Ca. femoratus</i> PS	Paracou	N 05°15.552 W 52°55.852	CHC only; library preparation failed
066-PAR	<i>Ca. femoratus</i> PS	Paracou	N 05°15.865 W 52°55.500	CHC only; removed due to skewed GC content
069-PAR	<i>Ca. femoratus</i> PS	Paracou	N 05°15.503 W 52°55.982	CHC only; library preparation failed
074-PAR	<i>Ca. femoratus</i> PS	Paracou	N 05°15.578 W 52°56.133	RNA + CHC
076-PAR	<i>Ca. femoratus</i> PS	Paracou	N 05°15.698 W 52°55.913	RNA + CHC

Table S2: Overview of all libraries with raw and trimmed read summary statistics. This table includes accession numbers for NCBI BioSample and NCBI SRA, the number of sequences before and after trimming, the GC content and the total number and percentage of reads that mapped to the de novo transcriptomes.

Sample ID	NCBI BioSample Accession	NCBI SRA Accession	Million Seqs before trimming	Million Seqs after trimming	GC content [%]	kallisto mapped reads [millions]	% mapped reads
63-PAR-Cr-FB2	SAMN11534817	SRS4702122	20,1	20,1	41%	18,822038	93,64%
64-PAR-Cr-FB2	SAMN11534818	SRS4702121	20,5	20,5	40%	18,755198	91,49%
75-PAR-Cr-FB1	SAMN11534819	SRS4702120	20,3	20,3	40%	17,924791	88,30%
231-PAT-Cr-FB1	SAMN11534820	SRS4702119	20,3	20,3	41%	18,640664	91,83%
240-PAT-Cr-FB2	SAMN11534821	SRS4702118	20,7	20,7	41%	18,697558	90,33%
242-PAT-Cr-FB2	SAMN11534822	SRS4702117	21	21	41%	19,155784	91,22%
59-PAR-Cr-FB2	SAMN11534823	SRS4702116	20,9	20,9	40%	19,500576	93,30%
69-PAR-Cr-FB2	SAMN11534824	SRS4702115	20,9	20,9	37%	19,449866	93,06%
74-PAR-Cr-FB2	SAMN11534825	SRS4702114	20	20	40%	18,419325	92,10%
76-PAR-Cr-FB2	SAMN11534826	SRS4702113	20,3	20,3	41%	18,471525	90,99%
218-PAT-Cr-FB1	SAMN11534827	SRS4702108	20,3	20,3	41%	19,199861	94,58%
219-PAT-Cr-FB1	SAMN11534828	SRS4702107	20,2	20,2	41%	19,058833	94,35%
222-PAT-Cr-FB2	SAMN11534829	SRS4702106	20,7	20,7	40%	18,949717	91,54%
230-PAT-Cr-FB2	SAMN11534830	SRS4702105	20,2	20,2	41%	18,76098	92,88%
232-PAT-Cr-FB1	SAMN11534831	SRS4702112	21	21	40%	19,317681	91,99%
51-PAR-Ca-FB1	SAMN11534832	SRS4702111	20,2	20,2	40%	18,825925	93,20%
59-PAR-Ca-FB1	SAMN11534833	SRS4702110	20,2	20,2	40%	18,830403	93,22%
77-PAR-Ca-FB2	SAMN11534834	SRS4702109	20,2	20,2	41%	18,982717	93,97%
218-PAT-Ca-FB1	SAMN11534835	SRS4702104	20,2	20,2	40%	18,851582	93,32%
232-PAT-Ca-FB2	SAMN11534836	SRS4702103	20,6	20,6	39%	19,273274	93,56%
240-PAT-Ca-FB2	SAMN11534837	SRS4702128	20,1	20,1	41%	18,804654	93,56%
242-PAT-Ca-FB1	SAMN11534838	SRS4702129	20,1	20,1	41%	18,73111	93,19%
53-PAR-Ca-FB1	SAMN11534839	SRS4702126	20	20	40%	18,983147	94,92%
55-PAR-Ca-FB1	SAMN11534840	SRS4702127	20,2	20,2	40%	19,153602	94,82%
63-PAR-Ca-FB2	SAMN11534841	SRS4702124	20,2	20,1	40%	18,953337	94,30%
74-PAR-Ca-FB2	SAMN11534842	SRS4702125	20,3	20,3	40%	19,244739	94,80%
76-PAR-Ca-FB1	SAMN11534843	SRS4702123	20,4	20,4	41%	18,873136	92,52%

Table S3: Sample combination for subassemblies. The table shows the CLC subassemblies and the samples used for those. Subassemblies were afterwards meta-assembled with MIRA.

Subassembly	Sample IDs
<i>Cr. levior</i> A #1	63-PAR-Cr-FB2 64-PAR-Cr-FB2 75-PAR-Cr-FB1
<i>Cr. levior</i> A #2	231-PAT-Cr-FB1 240-PAT-Cr-FB2 242-PAT-Cr-FB2
<i>Cr. levior</i> B #1	59-PAR-Cr-FB2 69-PAR-Cr-FB2 74-PAR-Cr-FB2
<i>Cr. levior</i> B #2	222-PAT-Cr-FB2 230-PAT-Cr-FB2 232-PAT-Cr-FB1
<i>Cr. levior</i> B #3	76-PAR-Cr-FB2 218-PAT-Cr-FB1 219-PAT-Cr-FB1
<i>Ca. femoratus</i> PAT #1	51-PAR-Ca-FB1 59-PAR-Ca-FB1 218-PAT-Ca-FB1 240-PAT-Ca-FB2
<i>Ca. femoratus</i> PAT #2	77-PAR-Ca-FB2 232-PAT-Ca-FB2 242-PAT-Ca-FB1
<i>Ca. femoratus</i> PS #1	53-PAR-Ca-FB1 55-PAR-Ca-FB1
<i>Ca. femoratus</i> PS #2	63-PAR-Ca-FB2 74-PAR-Ca-FB2 76-PAR-Ca-FB1

Table S4: Comparison of different assembly strategies using summary stats. Summary stats for assemblies of cryptic *Crematogaster* species (A) and cryptic *Camponotus* species (B) were created using *TransRate* (Smith-Unna *et al.* 2016).

A) *Crematogaster* cryptic species

assembly	N seqs	smallest	largest	N bases	Mean length	N under 200	N over 1k	N over 10k	N with orf	Mean orf percent	n90	n70	n50	n30	n10	GC	GC skew	AT skew	Cpg ratio	Bases N	Proportion N	Linguistic Complexity
Trinity	253516	301	26929	388095684	153085282	0	113918	674	98305	4244961	582	1607	2764	4169	6532	39429	4	-98	240775	0	0	23595
CLC Workbench only	118713	145	18660	72983341	61376398	631	14601	80	14587	5172291	256	424	930	2247	5097	39937	-152	-126	249175	191146	262	11095
CLC Workbench + MIRA	89286	80	38254	107916751	120565401	1617	25983	654	42553	9720813	396	1499	3063	5187	9319	39778	47	29	244502	451191	418	1781
final transcriptome (CLC + MIRA)	60185	299	38254	101019523	167848339	0	25983	654	42553	9829597	610	1794	3328	5418	9471	39741	66	4	244032	442237	438	23917

B) *Camponotus* cryptic species

assembly	N seqs	smallest	largest	N bases	Mean length	N under 200	N over 1k	N over 10k	N with orf	Mean orf percent	n90	n70	n50	n30	n10	GC	GC skew	AT skew	Cpg ratio	Bases N	Proportion N	Linguistic Complexity
Trinity	228541	301	24661	439174845	192164577	0	120007	1725	100386	4027705	784	2109	3513	5255	8186	38184	-36	-71	250764	0	0	26827
CLC Workbench only	89456	144	29106	64974968	72591701	193	13342	190	10991	4637211	274	528	1349	3136	6871	3759	-185	-94	263489	100686	155	12107
CLC Workbench + MIRA	70610	80	40738	87597661	123809691	1055	20323	636	33600	9634579	401	1552	3239	5489	9968	38001	-47	-26	257468	228586	261	17606
final transcriptome (CLC + MIRA)	48208	299	40738	82218009	170548475	0	20323	636	33600	97552	608	1869	3525	5715	10089	38031	-27	-16	256949	225206	274	23371

Table S5: List of genes known to be involved in CHC biosynthesis in *Drosophila*. We list the gene IDs, gene names, (putative) functions and source literature supporting its role in CHC biosynthesis.

Annotation ID in <i>Drosophila</i>	Gene name	Function	Literature
CG3524	<i>Fatty acid synthase 2</i>	Fatty acid synthase	(Chung <i>et al.</i> 2014)
CG17374	<i>Fatty acid synthase 3</i>	Fatty acid synthase	(Wicker-Thomas <i>et al.</i> 2015)
CG16905	<i>Elongase F</i>	Very long-chain fatty acid elongase	(Chertemps <i>et al.</i> 2007)
CG2781	<i>ELOVL fatty acid elongase</i>	Very long-chain fatty acid elongase	(Chiang <i>et al.</i> 2016)
CG6921	<i>bond</i>	Very long-chain fatty acid elongase of the Elov1 family	(Ng <i>et al.</i> 2015)
CG9458	<i>uncharacterized protein CG9458</i>	Putative very long-chain fatty acid elongase	(Dembeck <i>et al.</i> 2015)
CG18609	<i>uncharacterized protein CG18609</i>	Putative very long-chain fatty acid elongase	(Dembeck <i>et al.</i> 2015)
CG30008	<i>uncharacterized protein CG30008</i>	Putative very long-chain fatty acid elongase	(Dembeck <i>et al.</i> 2015)
CG5887	<i>Desaturase 1</i>	Delta(9)-Desaturase	(Dallerac <i>et al.</i> 2000; Labeur <i>et al.</i> 2002)
CG5925	<i>Desaturase 2</i>	Delta(9)-Desaturase	(Dallerac <i>et al.</i> 2000)
CG7923	<i>Desaturase F</i>	Delta(9)-Desaturase	(Chertemps <i>et al.</i> 2006)
CG13091	<i>uncharacterized protein CG13091</i>	Putative acetyl-CoA reductase	(Dembeck <i>et al.</i> 2015)
CG10097	<i>uncharacterized protein CG10097</i>	Putative acetyl-CoA reductase	(Dembeck <i>et al.</i> 2015)
CG17562	<i>uncharacterized protein CG17562</i>	Putative acetyl-CoA reductase	(Chiang <i>et al.</i> 2016)
CG18377	<i>Cyp49a1</i>	Putative cytochrome P450	(Dembeck <i>et al.</i> 2015)
CG11466	<i>Cyp9f2</i>	Putative cytochrome P450	(Dembeck <i>et al.</i> 2015)
CG9081	<i>Cyp4s3</i>	Putative cytochrome P450	(Dembeck <i>et al.</i> 2015)
CG3972	<i>Cyp4g1</i>	Cytochrome P450	(Qui <i>et al.</i> 2012)

References

- Chertemps, T., L. Duportets, C. Labeur, R. Ueda, K. Takahashi *et al.*, 2007 A female-biased expressed elongase involved in long-chain hydrocarbon biosynthesis and courtship behavior in *Drosophila melanogaster*. *Proc. Natl. Acad. Sci.* 104: 4273–4278.
- Chertemps, T., L. Duportets, C. Labeur, M. Ueyama, and C. Wicker-Thomas, 2006 A female-specific desaturase gene responsible for diene hydrocarbon biosynthesis and courtship behaviour in *Drosophila melanogaster*. *Insect Mol. Biol.* 15: 465–473.
- Chiang, Y. N., K. J. Tan, H. Chung, O. Lavrynenko, A. Shevchenko *et al.*, 2016 Steroid Hormone Signaling Is Essential for Pheromone Production and Oenocyte Survival. *PLoS Genet.* 12: e1006126.
- Chung, H., D. W. Loehlin, H. D. Dufour, K. Vaccaro, J. G. Millar *et al.*, 2014 A Single Gene Affects Both Ecological Divergence and Mate Choice in *Drosophila*. *Science* (80-.). 343: 1148–1151.
- Dallerac, R., C. Labeur, J. M. Jallon, D. C. Knipple, W. L. Roelofs *et al.*, 2000 A $\Delta 9$ desaturase gene with a different substrate specificity is responsible for the cuticular diene hydrocarbon polymorphism in *Drosophila melanogaster*. *Proc. Natl. Acad. Sci.* 97: 9449–9454.
- Dembeck, L. M., K. Böröczky, W. Huang, C. Schal, R. R. H. Anholt *et al.*, 2015 Genetic architecture of natural variation in cuticular hydrocarbon composition in *Drosophila melanogaster*. *Elife* 4: e09861.
- Labeur, C., R. Dallerac, and C. Wicker-Thomas, 2002 Involvement of *desat1* gene in the control of *Drosophila melanogaster* pheromone biosynthesis. *Genetica* 114: 269–274.
- Ng, W. C., J. S. R. Chin, K. J. Tan, and J. Y. Yew, 2015 The fatty acid elongase *Bond* is essential for *Drosophila* sex pheromone synthesis and male fertility. *Nat. Commun.* 6: 8263.
- Qui, Y., C. Tittiger, C. Wicker-Thomas, G. Le Goff, S. Young *et al.*, 2012 An insect-specific P450 oxidative decarbonylase for cuticular hydrocarbon biosynthesis. *Proc. Natl. Acad. Sci.* 109: 14858–14863.
- Smith-Unna, R., C. Bournnell, R. Patro, J. M. Hibberd, and S. Kelly, 2016 TransRate: Reference-free quality assessment of de novo transcriptome assemblies. *Genome Res.* 26: 1134–1144.
- Wicker-Thomas, C., D. Garrido, G. Bontonou, L. Napal, N. Mazuras *et al.*, 2015 Flexible origin of hydrocarbon/pheromone precursors in *Drosophila melanogaster*. *J. Lipid Res.* 56: 2094–2101.