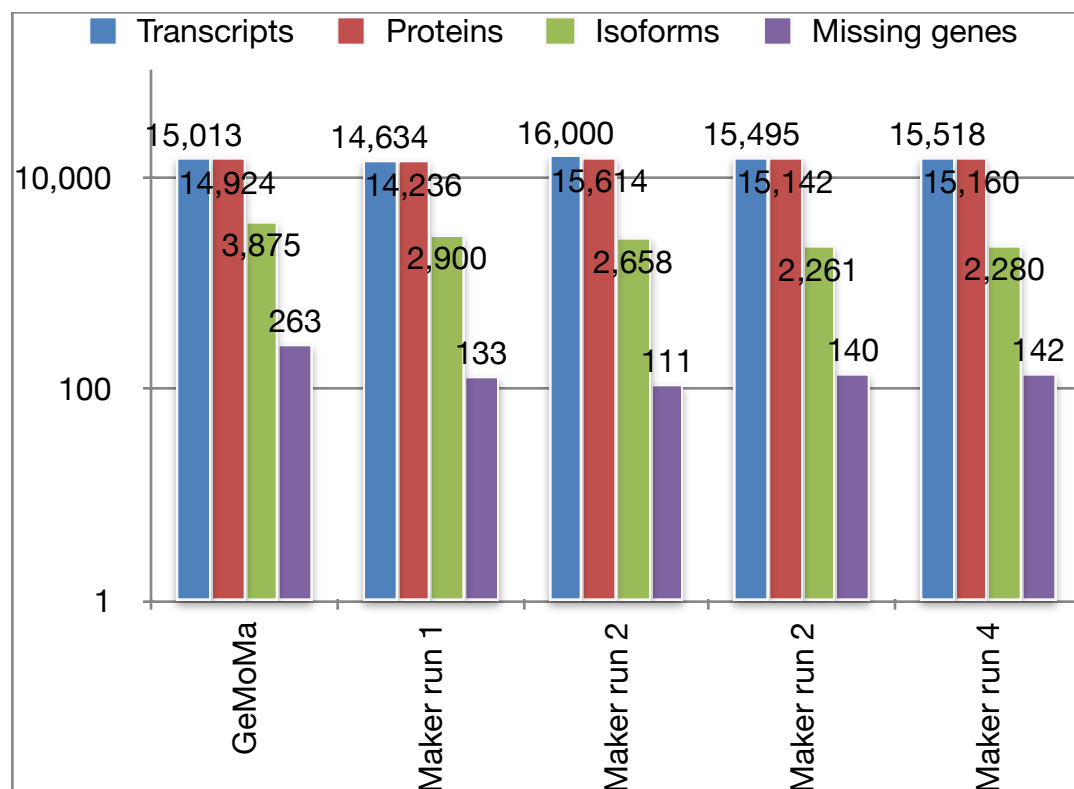


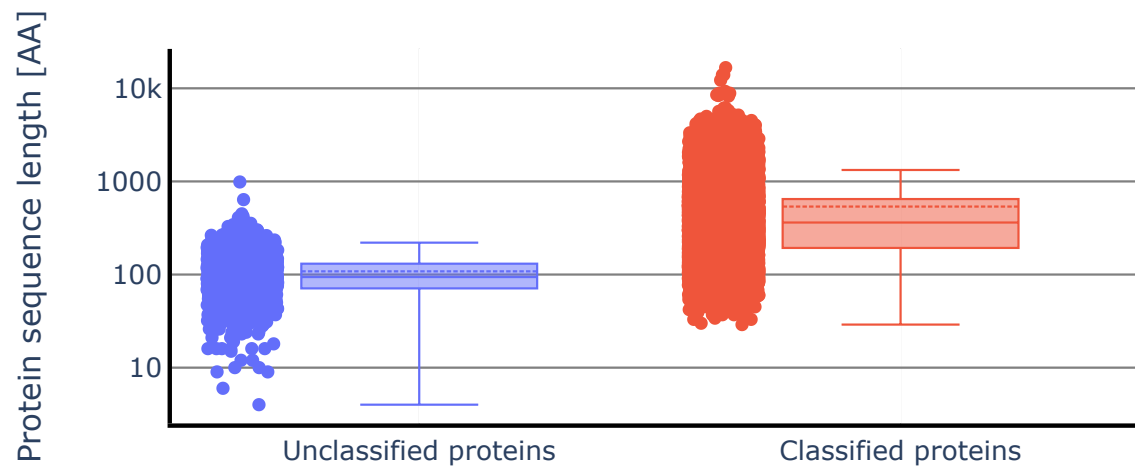
Genome Assembly and Annotation of the California Harvester Ant *Pogonomyrmex californicus* (Buckley, 1867)

Jonas Bohn^{*,1}, Reza Halabian^{*}, Lukas Schrader[†], Victoria Shabardina^{*,2}, Raphael Steffent[†], Yutaka Suzuki[‡], Ulrich R. Ernst[†], Jürgen Gadau^{†,3}, Wojciech Makalowski^{*,3}

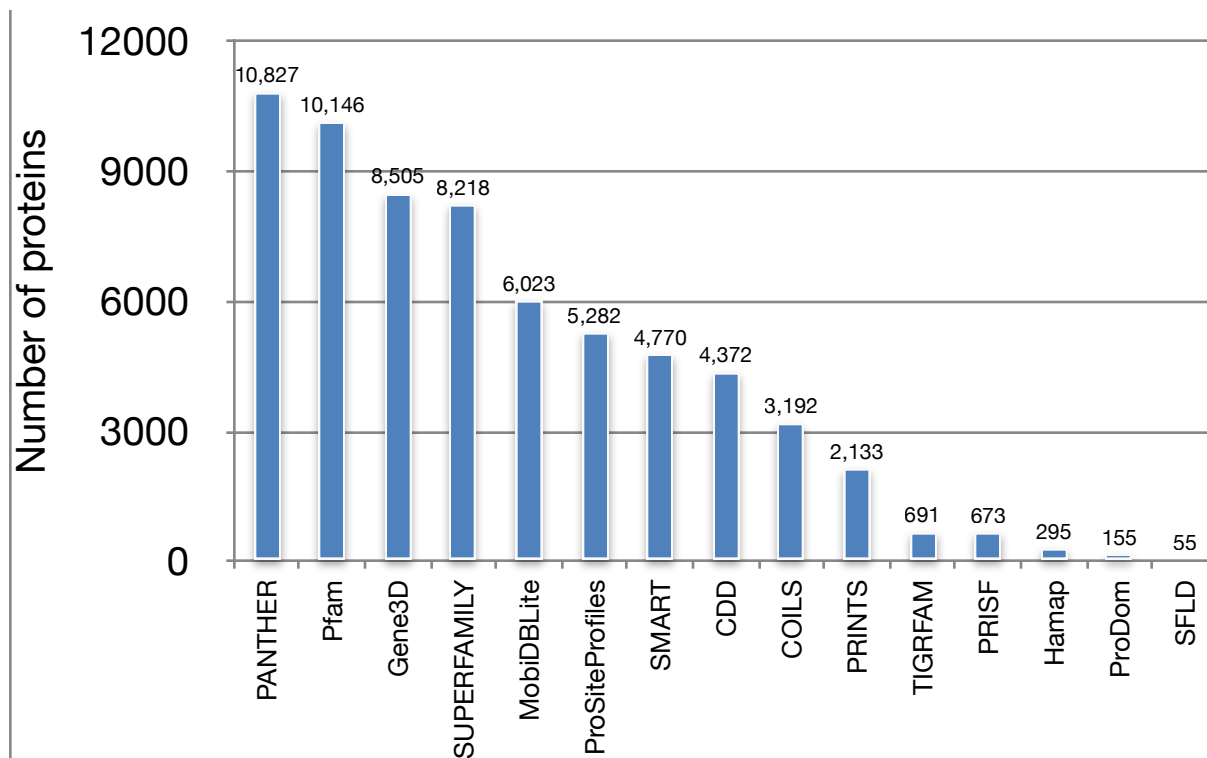
Supplementary materials



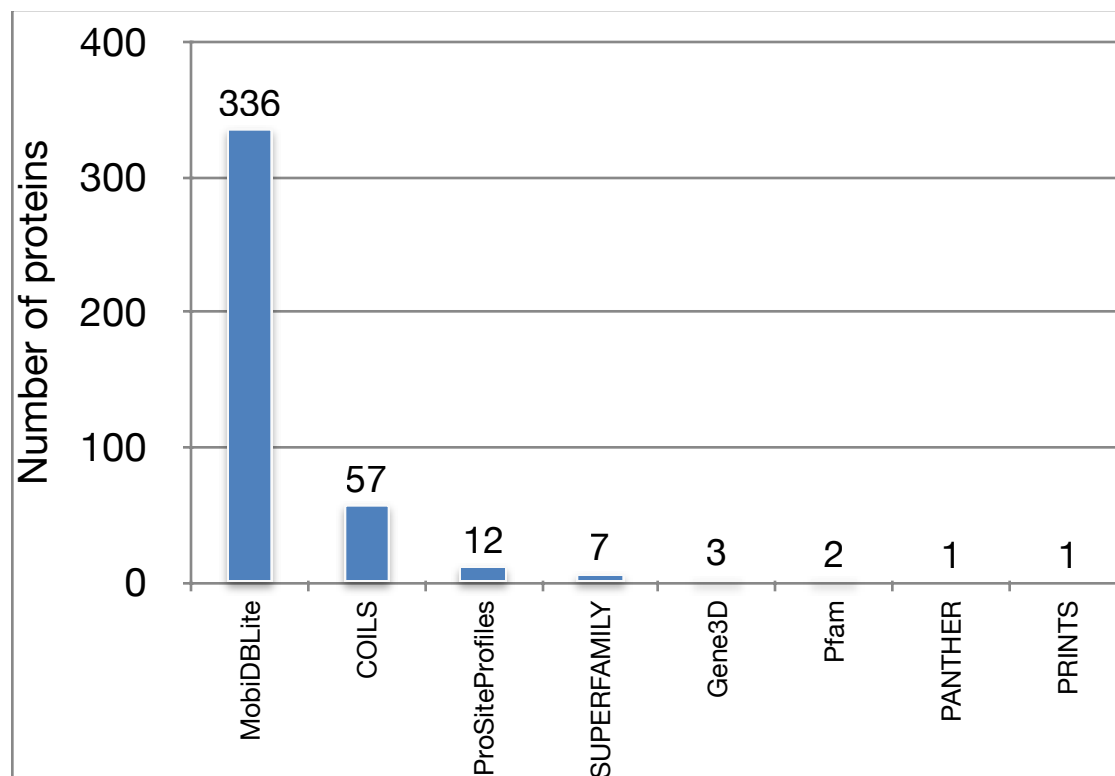
Supplementary Figure 1. Number of predictions and assessment of completeness through the annotation procedure. This analysis is based on unique/non-redundant sets of transcripts and proteins coming out of the annotation steps.



Supplementary Figure 2. Sequence length distribution of functionally classified and unclassified proteins. Please notice that the y axis is log scaled. The dotted line represents the mean and the solid line the median.



Supplementary Figure 3. Interproscan prediction distribution of classified proteins.



Supplementary Figure 4: Interproscan prediction distribution of non-classified proteins.

Supplementary Table 1. RefSeq accession numbers of assemblies and annotations from relative insects used for the gene annotation.

Species	RefSeq Accession	Assembly Name	Date of Access
<i>P. barbatus</i>	GCF_000187915.1	Pbar_UMD_V03	May 2018
<i>S. invicta</i>	GCF_000188075.1	Si_gnH	May 2018
<i>C. floridanus</i>	GCF_003227725.1	Cflo_v7.5	July 2018
<i>A. mellifera</i>	GCF_000002195.4	Amel_4.5	May 2018

Supplementary Table 2. Improving of the *P. californicus* assembly by polishing procedure.

	Assembly size (bp)	Number of Ns per 1 kb
Draft assembly	240,287,203	13.00
1st round of polishing	240,787,628	11.91
2nd round of polishing	241,014,132	11.64
3rd round of polishing	241,081,918	11.53

Supplementary Table 3. Number of OR gene models in the annotation of the *P. barbatus* genome available in NCBI as well as those predicted by the custom pipeline for both the *P. californicus* and *P. barbatus* genomes. Incomplete gene models are either missing N- and/or C-termini or, in case of the original annotation of *P. barbatus*, could not be identified as OR genes, but contain an OR domain. Homology based predictions do not contain an OR domain at all.

Species assembly	All OR gene models	Perfect OR gene models	Incomplete OR gene models	Homology based OR gene models
<i>P. barbatus</i> (manual annotation)	274	109	165	0
<i>P. californicus</i>	417	303	67	47
<i>P. barbatus</i>	454	342	43	69

Supplementary Table 4. Number of complete, incomplete, and homology based gene models predicted by the custom pipeline for both *P. californicus* and *P. barbatus* in the OR gene families.

OR gene family	Pcal complete	Pcal incomplete	Pcal domain missing	Pbar complete	Pbar incomplete	Pbar domain missing
9E	118	32	23	141	22	37
A	9	1	3	10	1	2
B	-	1	-	1	-	-
C	-	1	1	1	-	-
D	3	-	-	4	-	-
E	24	5	3	28	2	6
F	8	-	-	8	-	-
G	1	1	6	2	-	9
H	10	6	3	12	1	1
I	1	-	-	1	-	-
J	3	1	-	2	-	-
K	2	-	-	2	-	-
L	42	3	1	44	-	4
M	4	-	-	4	-	-
N	4	-	-	4	-	-
Orco	1	-	-	1	-	-
P	9	-	-	8	-	-
Q	1	-	-	1	-	-
R	1	1	-	1	-	1
S	1	-	-	1	-	-
T	1	3	-	6	-	-
U	18	7	1	19	13	3
V	40	5	3	39	4	4
W	1	-	-	1	-	-
X	-	-	1	-	-	1
XA	-	-	-	-	-	1
Z	1	-	2	1	-	-