1   The *Nicrophorus vespilloides* genome and methylome, a beetle with complex social behavior

2

3   Christopher B. Cunningham[1,*], Lexiang Ji[2], R. Axel W. Wiberg[3], Jennifer Shelton[4], Elizabeth C. McKinney[1],

4   Darren J. Parker[3], Richard B. Meagher[1], Kyle M. Benowitz[1], Eileen M. Roy-Zokan[1], Michael G. Ritchie[3],

5   Susan J. Brown[4], Robert J. Schmitz[1], Allen J. Moore[1,*]

6

7   Affiliations:

8

9   [1]Department of Genetics, University of Georgia, Athens, Georgia, USA

10

11   [2]Institute of Bioinformatics, University of Georgia, Athens, Georgia, USA

12

13   [3]Centre for Biological Diversity, School of Biology, University of St. Andrews, Fife, KY16 9TH, UK.

14

15   [4]Division of Biology & Bioinformatics Center & Arthropod Genomics Center, Kansas State University,

16   Manhattan, Kansas, USA

17

18   *Corresponding Authors: CB Cunningham (cbc83@uga.edu) and AJ Moore (ajmoore@uga.edu)

19

20   Running Title: Burying Beetle Genome

21

Abstract

23 Testing for conserved and novel mechanisms underlying phenotypic evolution requires a diversity of

24 genomes available for comparison spanning multiple independent lineages. For example, complex social

25 behavior in insects has been investigated primarily with eusocial lineages, nearly all of which are

26 Hymenoptera. If conserved genomic influences on sociality do exist, we need data from a wider range of taxa

27 that also vary in their levels of sociality. Here we present information on the genome of the subsocial beetle

28 *Nicrophorus vespilloides*, a species long used to investigate evolutionary questions of complex social

29 behavior. We used this genome to address two questions. First, does life history predict overlap in gene

30 models more strongly than phylogenetic groupings? Second, like other insects with highly developed social

31 behavior but unlike other beetles, does *N. vespilloides* have DNA methylation? We found the overlap in gene

32 models was similar between *N. vespilloides* and all other insect groups regardless of life history. Unlike

33 previous studies of beetles, we found strong evidence of DNA methylation, which allows this species to be

34 used to address questions about the potential role of methylation in social behavior. The addition of this

35 genome adds a coleopteran resource to answer questions about the evolution and mechanistic basis of

36 sociality.

37

Introduction

39 Understanding phenotypic evolution necessitates investigating both the ultimate and proximate influences on

40 traits; however, these investigations require the appropriate tools. Social behavior is a particularly thorny

41 phenotype to study because of its complexity, variation, and its multi-level integration across an organism

42 (Boake *et al.*, 2002). In addition, social behavior also often displays unusual evolutionary dynamics arising

43 from the genetic influences on interactions required for sociality (McGlothlin *et al.*, 2010). Although single

44 genes can influence behavior (Fischman *et al.*, 2011), social behavior is often multifaceted and can reflect a

45 complex genetic architecture (Walling *et al.*, 2008; Mikheyev and Linksvayer, 2015) including influences

46 from epigenetic mechanisms (Cardoso *et al.*, 2015). Genomes in particular are useful resources for

47 evolutionary questions of social behavior because they grant access to both broad scale and fine scale details

48 and mechanisms (Richards, 2015). For social behavior, while there are multiple Hymenopteran genomes

49 available to investigate fine scale detail, we lack sufficiently distantly related species to address broader

50 patterns. It is therefore important to develop genomic resources for organisms that are particularly

51 phenotypic models of social behavior but where genomic information is lacking.

52   The genomes of several social insect are now available, including eusocial species such as honey

53 bees (Elsik *et al.*, 2014), stingless bees (Kapheim *et al.*, 2015), several ant species (Gadau *et al.*, 2012), a

54 termite (Terrapon *et al.*, 2014), primitively eusocial species including bumble bees (Sadd *et al.*, 2015), and a

55 sweat bee (Kocher *et al.*, 2013). However, only one, the termite (Blattodea), is outside of the insect order

56 Hymenoptera. Although enormous progress has been made in identifying genes associated with the

57 behavioral division of labor and developmental shifts in social and other behavioral tasks (Mikheyev and

58 Linksvayer, 2015; Rehan and Toth, 2015), and the influence of epigenetic inheritance on developmental

59 plasticity and behavior (Glastad *et al.* 2015; Yan *et al.*, 2015), the generality of any mechanism underlying

60 social interactions requires information from insects reflecting other levels of sociality and from other orders.

61 Sociality at some level occurs in nearly all insect orders (Wilson, 1971; Costa, 2006), with eusociality

62 representing an extreme on a social continuum. Outside Hymenoptera there are many subsocial species that

63 have highly developed social behaviors, including parental care, but no division of labor (Costa, 2006). To

64 begin to address this gap, we assembled the genome of *Nicrophorus vespilloides*, a subsocial beetle that

65 serves as a behavioral model species for many types of complex social interactions, including elaborate and

66 advanced parental care with direct regurgitation of food to begging offspring (Walling *et al.* 2007), parent-

67 offspring conflict (Kilner and Hinde, 2012), sibling competition (Smiseth *et al.*, 2007), and adult competition

68 for resources (Hopwood *et al.*, 2013). By sequencing the genome of *N. vespilloides*, we were able to address

69 two questions: first, is the gene complement of *N. vespilloides* more reflective of phylogeny or life history?

70 Second, given methylation has been implicated in the success of eusocial species and facilitate plasticity,

71 could this mechanism play a role in *N. vespilloides* social plasticity as well? *Tribolium castaneum*, the model

72 beetle species, seems to lack DNA methylation (Zemach *et al.*, 2010). This has lead to the assumption that

73 methylation may be unimportant in beetles generally. However, methylation has been suggested to regulate

74 behavioral states in social insects (Glastad *et al.* 2015; Yan *et al.* 2015). *Nicrophorus vespilloides* is an

75 unusual beetle in that it is highly social, with extensive interactions between parents and offspring, but males

76 in the presence of females do not care for offspring or show the same levels of gene expression as caring

77 parents (Parker *et al.*, in review). There is a rapid transition between behavioral states if the female parent is

78 removed (Smiseth *et al.* 2005), with extensive changes in gene expression in the male (Parker *et al.*, in

79 review). Given this, we hypothesized the presence of DNA methylation in *N. vespilloides*, which could

80 provide a mechanism for this rapid behavioral transition.

81 Burying beetles (*Nicrophorus spp.*) are a group of about 70 species that have a long history as a

82 model species used to investigate the evolution of complex social behaviors (Eggert and Müller, 1997; Scott,

83 1998; Sikes and Venables, 2013). Burying beetles use vertebrate carcasses as food for their offspring, but go

84 well beyond simple forms of parental care such as providing food and protection, and display extensive and

85 complex family interactions (Eggert and Müller, 1997; Scott 1998; Trumbo 2012). Once suitable carrion is

86 located, these beetles remove hair, feathers, or scales, form the carcass into a ball, and spread anti-microbial

87 excretions over its surface to retard decomposition by microbial growth. During the period of carcass

88 preparation, the female deposits eggs in the soil around the carcass and upon hatching larvae crawl to the

89   carcass and enter a small cavity opened by the parents in the carcass. In the next phase, there are extensive

90   and direct parent-offspring interactions involving begging by offspring and direct regurgitation of digested

91   carrion by parents to begging offspring. Parents also pre-digest carrion and continue to process the carcass by

92   mastication and excreting digestive enzymes into the larval cavity to assist larvae with easy food acquisition.

93   Parents also continue to maintain the carcass against microbial growth and interspecific competitors (e.g., fly

94   larvae). Both parents can be present, and there is communication between parents as well as between parents

95   and offspring. The most extensively studied burying beetle, *N. vespilloides*, has proven an excellent model

96   for investigating the ecology and evolution of social interactions between family members (Eggert and

97   Müller, 1997; Scott, 1998; Trumbo, 2012; fig. 1). Although parental care is essential, especially in the first

98   24 hours of larval life (Eggert *et al.*, 1998; Smiseth *et al.*, 2003), care in this species can be uniparental,

99   either male or female, or biparental. The parental care expressed is identical in all three conditions, and all

100 forms of care are equivalently beneficial for offspring (Eggert, 1990; Parker *et al.*, *in review*). Although

101 parents attend the carcass for 95 h after larval arrival (Parker *et al.*, *in review*), there is variation among

102 parents in how much time they allocate to care and its different components, and these differences are

103 heritable in *N. vespilloides* (Walling *et al.*, 2008).

104   Here, we report a genome assembly of *N. vespilloides* and use this to investigate hypotheses

105 regarding evolution associated with social behavior. Our assembly integrates high-throughput short reads,

106 long reads, and a genome map providing sequence for greater than 90% of the predicted genome size. We

107 annotated 13,526 protein-coding genes and compared these genes to social species, another beetle, and other

108 insects that use vertebrate carcasses as food but lack sociality. The overlap of shared number of orthologs

109 was similar between *N. vespilloides* and all other insect groups regardless of life history. We then tested if *N.*

110 *vespilloides* has DNA methylation by looking for the enzymes responsible for DNA methylation and further

111 by using whole genome bisulfite sequencing with the hypothesis that *N. vespilloides* would lack DNA

112 methylation as does *T. castaneum* (Zemach *et al.*, 2010). Unlike previous studies of beetles, we found

113 evidence of DNA methylation. This genome adds the first coleopteran resource to answer questions about the

114 evolution and mechanistic basis of complex social behavior.

115

116 **Results**

117 **Genome Sequencing and Assembly.** We assembled the genome of *N. vespilloides* by integrating evidence

118 from Illumina short reads, Pacific Bioscience (PacBio) continuous long reads, and a BioNano Genomics

119 genome map (supplementary table S1). We assembled 195.3 Mb of the *N. vespilloides* genome, which is

120 95.7% of its predicted size (supplementary fig. S1). The draft genome is contained within 5,858 contigs with

121 an N50 of 102.1 kb and further into 4,664 scaffolds with an N50 of 122.4 kb (Longest scaffold: 1.80 Mb;

122   table 1). The GC content is 32%, consistent with two other beetle genomes, *T. castaneum* at 33% (Tribolium

123   Genome Sequencing Consortium, 2008) and *Dendroctonus ponderosae* at 36% (Keeling *et al.*, 2013).

124   We assessed how well the protein-coding portion of the genome was assembled using the CEGMA

125   and BUSCO pipelines. Our genome contained 247 complete orthologs (99.6%) and 248 partial orthologs

126   (100%) of the CEGMA proteins. Of the 2,827 *T. castaneum* BUSCO proteins, our genome contained 2,737

127   (96.8%) as single-copy orthologs and 86 (3.1%) as multi-copy orthologs. We also mapped the RNA-Seq data

128   used for annotation back to the genome to assess transcriptome coverage. There was an 89.7% mapping rate.

129

130   **Genome Annotation.** We used Maker2 to annotate the protein-coding portion of the genome by integrating

131   *ab initio*, protein homology, and species-specific RNA-Seq evidence into consensus gene models. We

132   obtained 13,526 predicted gene models. The gene models had an average protein length of 466.7 amino acids

133   and 6.3 exons. Maker2 also predicted 5´ untranslated regions (UTR's) for 5,813 genes (mean: 512 bp) and 3'

134   UTR's for 4,549 genes (mean: 980 bp).

135   We were able to functionally annotate 11,585 gene models (85.6%) against UniProtKB with

136   BLASTp. Restricted to species that had five or more best matches against *N. vespilloides* (encompassing

137   97.8% of the annotated gene models), the annotated gene models overwhelmingly returned the strongest

138   similarity to other Coleoptera (fig. 2, supplementary table S2; top three species- *T. castaneum*: 6,969, *D.*

139   *ponderosae*: 1,368, *Anoplophora glabripennis*: 743; Coleopteran total: 9,210 (79.5%)). Arthropods were the

140   strongest similarity matches for 11,305 (99.7%) gene models (fig. 2). We were also able to identify at least

141   one protein domain in 86.1% of the genes using InterProScan5. Searches against the Pfam database found

142   9,467 domains from 3,932 unique families. We were also able to assign at least one Gene Ontology (GO)

143   term to 7,492 genes (55.4%). Additionally, we were able to associate KEGG orthology terms with 44.8% of

144   the genes.

145   Our *de novo* repeat analysis found that 12.85% of the draft genome is composed of repetitive

146   elements. The top three classifications of repeats were: unclassified repetitive elements (6.13%), DNA

147   elements (3.35%), and simple repeats (2.24%). The overall repeat content is lower than that reported for

148   beetles *T. castaneum* (Tribolium Genome Sequencing Consortium, 2008) and *D. ponderosae* (Keeling *et al.*,

149   2013), but higher than the honey bee (Elsik *et al.*, 2014) and the red harvester ant (Smith *et al.*, 2012), all of

150   which have genomes that are of comparable size to *N. vespilloides*. Additionally, when we provided our

151   repeat library to RepeatMasker to mask the *T. castaneum* genome only 1.65% was masked, an outcome

152   consistent when the repeat library of *D. ponderosae* was used for the same task (0.15% of *T. castaneum*

153   masked; Keeling *et al.*, 2013; supplementary table S3).

154

155   **Orthology of Gene Models.** We used OrthoMCL, which clusters proteins based on a reciprocal best BLAST

156   hit strategy, to assign orthology of the *N. vespilloides* proteome against five other insect proteomes chosen

157   either because they are genomic models (*T. castaneum, Drosophila melanogaster*) or because of their life

158   history (*Apis mellifera, Musca domestica, Nasonia vitripennis*). Our analysis produced 11,929 orthologous

159   groupings with representatives from at least two different lineages. There were 4,928 orthologs groupings

160   that contained at least one protein from each species. Of these, 3,734 groups were single copy orthologs

161   among the six insects. There were 153 groupings containing 532 proteins that had proteins from *N.*

162   *vespilloides* only. The beetles, *N. vespilloides* and *T. castaneum*, were represented in 7,827 groupings and

163   716 groupings were exclusive to beetles (650 were single copy ortholog groupings). We then made two

164   specific comparisons of the proteomes of *N. vespilloides*, *T. castaneum*, *A. mellifera*, *N. vitripennis*, *D.*

165   *melanogaster,* and *M. domestica* (fig 3). We first compared *N. vespilloides* to a social (*A. mellifera*) and non-

166   social (*N. vitripennis*) Hymenopteran, including *T. castaneum* as a non-social beetle (fig. 3A). There was no

167   increased similarity in the number of protein families between *N. vespilloides* and the Hymenoptera than

168   there was between *T. castaneum* and the Hymenoptera. We next compared the proteome of *N. vespilloides* to

169   a Diptera that uses rotting carcasses for reproduction, *M. domestica* and again included *T. castaneum* and

170   another dipteran, *D. melanogaster* (fig 3B). Again, there were no obvious differences in overlap between *N.*

171   *vespilloides* and the Diptera that reproduces on carrion and *T. castaneum* and *D. melanogaster*. Thus, the

172   predicted proteins of *N. vespilloides* appear to be as similar to the other insects as is another beetle, *T.*

173   *castaneum* and there is no obvious specialization associated with sociality or reproductive mode.

174

175   **Gene Family Expansion and Contraction**. To investigate if there had been any gene family expansions or

176   contractions in *N. vespilloides*, we analyzed the results of the OrthoMCL analysis with CAFÉ. There were

177   269 orthology groupings (or gene families) that showed significant expansion or contraction between the six

178   insect species compared at $P < 0.0001$. Of these groupings 12 showed significant differences within the *N.*

179   *vespilloides* lineage. There were 8 expansions and 4 contractions (supplementary file S1). The expansions

180   were mostly families of uncharacterized proteins (7/8), while the last family was a chymotrypsin protease.

181   There was not an enrichment of any GO term from the expanded gene families. The contracted families had

182   highest similarity to an esterase, a transpoase, a cytochrome P450, and an uncharacterized protein in *T.*

183   *castaneum*. Some of these are also differentially expressed during caring (Parker *et al.*, in review).

184

185   **Selection Analysis.** Signatures of selection on the protein coding genes of *N. vespilloides* was investigated

186   by comparing the *dN/dS* (ω) ratio to *T. castaneum* and *D. ponderosae* for the 5,584 one-to-one orthologs we

187   detected between these lineages. Twenty five genes showed signs of differential divergent selection in the *N.*

188   *vespilloides* lineage after our filtering criteria were applied (BH false discovery rate = 0.05 and removal of

6

189   genes showing *dN, dS,* or *ω* >10). The distributions of these estimates are shown in fig 4**.** Two genes show

190   evidence of positive selection *ω* > 1. Ephrin-B2 (*efn-b2*; *ω* = 1.45) and NK Homeobox (HOX) 7 (*nk7*; *ω* =

191   2.16). *efn-b2* also has a *ω* > 1 in the other lineages (*ω* = 1.5) while *nk7* shows evidence of strong

192   conservation in the *T. castaneum* and *D. ponderosae* lineages. The median estimates of *dS, dN* and *ω* were

193   higher in the *N. vespilloides* lineage (*N. vespilloides*- 0.0489, *T. castaneum*- 0.0487 and *D. ponderosae*-

194   0.0487), although not statistically significantly different.

195

196   **DNA Methylation.** We used two approaches to investigate if the *N. vespilloides* genome has active DNA

197   methylation. First, we looked for the enzymes responsible for methylation in animals (Dnmt1, Dnmt2, and

198   Dnmt3) to determine if the machinery was present for the establishment and maintenance of DNA

199   methylation. Second, we generated single-base resolution maps of DNA methylation using whole genome

200   bisulfite sequencing.

201   Single copies of all three DNA methyltransferases were in the *N. vespilloides* genome; *T. castaneum*

202   contains only Dnmt1 and Dnmt2 (Kim *et al.*, 2010). The methyltransferases clustered with their putative

203   orthologs (fig. 5A). Next, using MethylC-Seq we found direct evidence for DNA cytosine methylation in *N.*

204   *vespilloides* and no evidence for DNA methylation in *T. castaneum*, supporting previous reports on the latter

205   (fig. 5B; Zemach *et al.*, 2010). Methylation (5´-methlycytosine) in *N. vespilloides* was found within a CpG

206   context exclusively (fig. 5C). A small proportion (1.87%) of CpH (H= A,T, or C) was found during the first

207   analysis; however, further analysis of the originally identified CpH methylated sites revealed that >98% of

208   them were artifacts of segregating SNPs. Therefore, only strong evidence was found for CpG methylation in

209   the genome. Methylated cytosines in *N. vespilloides* exhibited several typical insect patterns by being either

210   reliably methylated or not (fig. 5D) and exhibited a high level of symmetrical methylation on opposing DNA

211   strands (fig. 5E), a pattern typical of animals. The genome-wide pattern of DNA methylation observed for *N.*

212   *vespilloides* is also similar to other insects. Most prominently, the majority of methylation was found in the

213   exons (62.55 ± 0.26% of the observed methylation) and much lower levels were found in introns (10.29 ±

214   0.12% of the observed methylation; fig. 5F). All three biological replicates are quantitatively similar in their

215   distribution of methylated CpG's over gene elements (supplementary table S6). We grouped *N. vespilloides*

216   genes as methylated or non-methylated by comparing the level of methylation of an individual gene to the

217   average level of gene methylation found across all genes. We found 3,298 genes that were methylated

218   significantly higher than the null expectation (fig. 5G; supplementary file S2). Following this, we performed

219   a GO enrichment analysis on the GO terms associated with the methylated gene set. We found that

220   translation (GO:0006412), translation factor activity/nucleic acid binding (GO:0008135), and RNA binding

221   (GO:0003723) were significantly enriched molecular function GO terms. Cellular macromolecule metabolic

222   process (GO:0044260), gene expression (GO:0010467), and macromolecule biosynthesis process

223   (GO:043170) were the three most enriched biological process GO terms (see also supplementary table S7).

224   At the level of individual genes, methylation was highest in the most exons of a gene (Fig. 5H). Methylation

225   was also observed in the 5´- and 3´-UTR's, with the typical steep decrease in methylation observed at the

226   translational start site. We also observed methylation in the "promoter" region 1kb upstream from the first

227   annotated gene element. Methylation peaks beginning at the second exon, although this is not a robust trend

228   as methylation levels decrease to the same level of the first exon by the end of the second exon. Transposable

229   elements were methylated to the same level as genomic intergenic background levels (3 vs. 5%,

230   respectively).

231

232   Discussion

233   The ability to detect conserved and novel molecular mechanisms that influence social behavior requires

234   genomic resources from species across different lineages that vary in their level of sociality. Here, we report

235   the draft genome of *N. vespilloides*, a subsocial beetle from the Silphidae. In assessing the genetic changes

236   associated with the evolution of social behavior in insects, the *N. vespilloides* genome provides a useful line

237   of independent evolution, offering data from outside the Hymenoptera, which diverged from Coleoptera

238   ~350 Mya (Wiegmann *et al.*, 2009) and at a level between solitary and eusocial. *Nicorphorus vespilloides*

239   has sophisticated and complex parental care (Eggert and Müller, 1997; Scott, 1998; Trumbo, 2012). The

240   highly developed social interactions between parents and offspring place this beetle at the level of

241   "subsocial" on the evolutionary spectrum of social species (Wilson 1971; Costa 2006). Thus, we provide

242   investigators interested in the genomic and molecular signatures of social interactions, parent-offspring

243   conflict, social tolerance, mate choice, and mate cooperation with an experimentally tractable and

244   evolutionarily divergent model to use in comparative studies. In addition, unlike Hymenoptera and the

245   eusocial termites, where parental care is associated with developmental transitions to behavioral tasks, in *N.*

246   *vespilloides* parental care is reversible and repeatable as individuals switch between solitary to social to

247   solitary again across breeding cycles. Moreover, in *N. vespilloides*, both males and females express care

248   behavior, providing a decoupling of reproductive physiology and social interactions (Roy-Zokan *et al.,*

249   2015). Finally, *N. vespilloides* represents an evolutionarily divergent beetle to the other available coleopteran

250   genomes (~240 Mya from both *T. castaneum* and *D. ponderosae;* Hunt *et al.*, 2007).

251   We successfully assembled the *N. vespilloides* genome using Illumina short reads, PacBio

252   continuous long reads, and a BioNano Genomics genome map. Our assembly quality compares favorably

253   with other recently published insect genomes; especially considering our organism is outbred (Richards and

254   Murali, 2015). We found that our genome is similar to other recently sequenced insect genomes, with a

255   comparable number of genes and percentage of genes having a functional annotation (Kim *et al.*, 2010;

256   Wurm *et al.*, 2012; Keeling *et al.*, 2013; Oxley *et al.*, 2014; Wang *et al.*, 2014). Our orthology analysis

257 showed that *N. vespilloides* was as, but no more, similar to a social Hymenoptera or to Diptera that exist in

258 rotting flesh than the asocial beetle *T. castaneum* was to either pair with respect to the number of shared gene

259 families.

260   Very few of the *N. vespilloides* genes we examined showed evidence of differential rates of

261 sequence evolution compared to the *T. castaneum* and *D. ponderosae* lineages. Among the genes that did

262 show differential *dN/dS* ratios, the majority showed constrained evolution. We found only two genes with

263 evidence of positive selection. NK Homeobox (HOX) 7 had an elevated ω in the *N. vespilloides* lineage but

264 is highly conserved in the other lineages. Ephrin-B2 had an elevated ω in all lineages but it is slightly lower

265 in the *N. vespilloides* lineage. Both of these genes are involved in patterning (Dönitz *et al.*, 2015; Dos Santos

266 *et al.*, 2015). Overall, the genes compared show a high degree of conservation. One limitation of this analysis

267 is the approximately 240 Mya of evolutionary distance between *N. vespilloides* and *T. castaneum*. Moving

268 forward, it would be interesting to see how robust these results are to other types of analyses of molecular

269 evolution and as more beetle species over a range of phylogenetic distances are available for comparison.

270   Beetles are often described as lacking DNA methylation, based on *T. castaneum* (Glastad *et al*, 2015;

271 Yan *et al.* 2015). We have direct evidence for DNA methylation in the *N. vespilloides* genome. It is

272 intriguing that a social beetle has this trait, which suggests the potential for an overlap in epigenetic

273 mechanisms influencing social traits. Regardless, our works shows that lacking methylation is not a general

274 feature of Coleoptera. In fact, methylation in *N. vespilloides* looks very similar to most other insects with

275 active systems of methylation. *Nicrophorus vespilloides* has DNA methylation that is restricted to CpG sites

276 at levels similar to honey bees (Lyko *et al.*, 2010) and the jewel wasp *Nasonia vitripennis* (Wang *et al.*,

277 2013), the ants *Camponotus floridanus* and *Harpegnathos saltator* (Bonasio *et al.*, 2012), a grasshopper

278 *Schistocerca gregaria* (Falckenhayn *et al.*, 2013), a locust *Locusta migratoria* (Wang *et al.*, 2014), and the

279 silk moth *Bombyx mori* (Xiang *et al.*, 2010). Methylation is concentrated within the exons of genes as seen

280 with honey bees (Lyko *et al.*, 2010), ants (Bonasio *et al.*, 2012), the jewel wasp (Wang, *et al.*, 2013), but

281 different from a locust (Wang *et al.*, 2014) and the silk moth (Bonasio *et al.*, 2012). Methylation was also

282 found in the UTR's, a pattern also reported in *C. floridanus* and *H. saltator* (Bonasio *et al.*, 2012).

283 Methylation peaks at the beginning of the second exon, a pattern seen in ants (Bonasio *et al.*, 2012) and the

284 jewel wasp (Wang *et al.*, 2013). Differential DNA methylation has been implicated in the transition between

285 behavioral states in social insects (Lyko *et al.*, 2010; Bonasio *et al.*, 2012; Herb *et al.*, 2012; Terrapon *et al.*,

286 2014). Because *N. vespilloides* demonstrates dramatic and reversible switches in behavioral states across a

287 breeding cycle, and can have multiple breeding cycles, we hypothesize that DNA methylation is an

288 epigenetic mechanism that influences these behavioral transitions and stability.

289   Studies of the genetic basis and evolution of complex social behavior have focused on specific

290 genes, with conflicting results. However, these are mostly focused on age-based division of labor in the

291   eusocial Hymenoptera (Mikheyev and Linksvayer 2015; Rehan and Toth 2015). The addition of the *N.*

292   *vespilloides* genome allows us to expand beyond hymenopteran-specific aspects of social behavior, and

293   allows us to begin to address broader categories of social traits. Although there are numerous aspects of the

294   life history of burying beetles that make them unique (Eggert and Müller, 1997; Scott 1998), here we have

295   emphasized the value of using *N. vespilloides* as a model for studying family social interactions and social

296   evolution. These beetles are particularly suited for questions of parental care because the phenotype is robust

297   and readily measured, contains diverse sub-behaviors that are reliably observed and scored, can vary between

298   males and females in the context in which it is expressed, and is highly replicable (Walling *et al.* 2008). With

299   the addition of the *N. vespilloides* genome alongside Hymenoptera and a eusocial termite, we now have a

300   taxonomically diverse arsenal of phenotypically overlapping organisms to look for phylogenetically

301   independent genomic mechanisms and signatures of evolution, conservation, and novelty.

302

303   **Materials and Methods**

304   **Animals Samples.** All *Nicrophorus vespilloides* used in this research were obtained from an outbred colony

305   maintained at the University of Georgia under laboratory conditions for this species (see Cunningham *et al.*,

306   2014 for a full description of conditions).

307

308   **Genome Size Estimation, Sequencing, Assembly, and Quality Control.** We used flow cytometry with

309   propidium iodide staining to estimate the genome size of *N. vespilloides* using *T. castaneum* as a standard.

310   Nuclei from insect heads and whole insects, respectively, were prepared as described in Yu *et al.* (2015),

311   stained as in Hare and Johnston (2011), and analyzed with a CyAn Flow Cytometer (Beckman Coulter, Brea,

312   CA) at the UGA's Center for Tropical and Emerging Global Diseases Flow Cytometry Core Facility. Data

313   was processed with FlowJo software (Treestar, Inc., Ashland, Oregon).

314   Genomic DNA was extracted from a single larva derived from a single sibling-sibling mating using a

315   SDS-lysis buffer and a phenol-chloroform extraction. A 275 bp Illumina (San Diego, CA, USA) TruSeq

316   library was prepared and run on one lane of an Illumina HiSeq 2000 using a paired-end (2 x 100 bp)

317   sequencing protocol at the HudsonAlpha Genome Sequencing Center (Huntsville, AL, USA).

318   We used FastQC (v0.11.2; Babraham Institute; default settings) to create summary statistics and to

319   identify possible adapter contamination of raw Illumina paired end reads. No adapter contamination was

320   reported, a result supported by analysis with CutAdapt (v1.2.1; Martin, 2011), which only found evidence for

321   adapters in less than 0.01% of the raw reads. Because sequencing library construction can generate inserts of

322   genomic DNA that are less than twice the average read length, overlapping paired-end reads were first

323   merged using FLASH (v1.2.4; Magoc and Salzberg, 2011; default settings, insert size: 278 bp with SD of 53

324   bp (estimate from Platanus scaffolding step)). Quality control was performed with PrinSeq (Schmieder and

10

325   Edwards, 2011a). Reads were required to have a mean overall Phred quality score of $\geq$ 25, read ends were

326   trimmed to >20 Phred quality score, a minimum length of 90 bp and a maximum length of 99 bp were

327   allowed, and reads were allowed only one unidentified (N) nucleotide per read.

328   To obtain Pacific Bioscience (PacBio; Menlo Park, CA, USA) continuous long read, we extracted

329   genomic DNA using the same phenol-chloroform extraction as used to extract gDNA for the Illumina

330   sequencing from a brother/sister pair of adult beetles that had been inbred for 6 generations. A 14.4kb long

331   insert PacBio library was prepared by the University of Maryland Institute for Genomic Sciences. This

332   library was sequenced with 22 PacBio's RS II P5-C4 SMRT cells to generate continuous long read (CLR's)

333   to scaffold our assembly with to increase long-range connectivity of the genome. PacBio reads greater than

334   6300 bp (36.4x coverage) were error corrected with the PBcR pipeline (Koren *et al.*, 2012) using 49x

335   coverage of the quality-controlled Illumina reads with default settings, which after error correction and

336   assembly produced an estimated 20.9x coverage of CLRs.

337   To increase the long-range scaffolding (i.e., super-scaffold) of our draft genome, we generated a

338   BioNano Genomics (San Diego, CA, USA) genome map. High molecular weight (HMW) genomic DNA

339   was extracted from a single pupa as previously described (Shelton et al., 2015). HMW gDNA was nicked

340   with nicking restriction digest by BspQI and BbvCI restriction enzyme that had been converted to nickases

341   (New England Biolabs, Ipswich, MA, USA). Restriction sites were labeled with fluorescent nucleotides and

342   imaged on the Irys system (BioNano Genomics) according to the manufacturer's instructions.

343   All Illumina read passing quality control were used as input for the Platanus assembler (v1.2.1;

344   Kajitani *et al.*, 2014). First, reads were assembled into contigs using Platanus's assemble protocol (non-

345   default settings: -s 3 -u 0.2 -d 0.3 -m 128). Next, contigs were scaffolded using Platanus's scaffold protocol

346   (non-default settings: -u 0.2). This step was iterated a total of five times using the same settings to extend the

347   scaffold as much as possible with the Illumina reads. Gaps in the assembly were filled using Platanus's

348   gap_close protocol with default settings. This step was iterated twice. Only contigs/scaffolds 1kb or greater

349   in length were used for further analysis and assembly.

350   PacBio reads were used to gap fill and scaffold the Platanus assembly with PBJelly2 (v14.9.9;

351   English *et al.*, 2012) using default settings and the error corrected PacBio CLR's reads.

352   A genome map created from BioNano Genomic single molecule maps was used to super-scaffold the

353   Platanus/PBJelly assembly (Shelton *et al.*, 2015). Briefly, the images were assembled into a consensus map

354   based on the labeling pattern of each molecule imaged. These *in silico* maps, with a cumulative length of

355   133.7 Mb, were compared to the predicted labeling pattern of the Platanus/PBJelly that passed a quality filter

356   (length > 150kb and number of labels >= 8) to further scaffold and orient the Platanus/PBJelly assembly.

357   DeconSeq (v0.4.2; Schmieder and Edwards, 2011b) was used to assess our draft assembly for

358   possible contamination. Besides the 1,126 bacterial species included in the distribution, we also updated the

11

359   human genome sequence (h37) and added the genomes of *Caenorhabditis elegans*, *Ralstonia pickettii*,

360   *Yarrowia lipolytica*. *C. elegans* was included because it is the closest genome available to the nematode

361   symbiont of *N. vespilloides*, *Rhabditis stammeri* (Richter, 1993). *R. pickettii* and *Y. lipolytica* were included

362   because they were two species that showed up when the RNA-Seq experiment was assessed for possible

363   contamination (Parker *et al.*, in review). *Tribolium castaneum* was used as a retention database. Only one

364   contig was flagged and removed during our contamination search; belonging to *Morganella morganii*, a

365   common bacterium found in vertebrate intestinal tracts.

366   Genome assembly quality and completeness were assessed with multiple benchmark datasets. First,

367   the CEGMA analysis pipeline (v2.4.010312; Parra *et al.*, 2009) was used to assess the completeness of 248

368   ultra conserved eukaryotic genes within our assembly. Next, we used the *T. castaneum* set of Benchmarking

369   sets of Universal Single-Copy Orthologs (BUSCO's; 2,787 genes) to further assess the assembly

370   completeness (Waterhouse *et al.*, 2013). We also mapped the RNA-Seq reads back to the assembly to

371   estimate coverage of the transcriptome of our assembly using the TopHat (v2.0.13) pipeline with Bowtie2

372   (v2.2.3) as the read aligner.

373

374   **Genome Annotation.** To begin genome annotation, we first generated a *de novo* library of repeats using

375   Repeat-Modeler (v1.0.8) that integrates three separate repeat finder programs; RECON (v1.08; Bao and

376   Eddy, 2002), RepeatScout (v1.05; Price *et al.*, 2005), and TRF (v4.07b; Benson, 1999) with default

377   parameters. Because some gene fragments, especially low-complexity motifs, might be captured in the repeat

378   analysis, we used BLASTx to remove any matches to *T. castaneum* proteins in the UniProtKB database

379   (Wang *et al.*, 2008; Jiang, 2014). The repeat analysis of the *T. castaneum*'s genome was done with

380   RepeatMasker (v4.0.5; Smit *et al.*, 2015) using default settings.

381   We annotated putative protein coding genes using the Maker2 annotation pipeline (v2.31.7; Holt and

382   Yandell. 2011) using an iterative process. After masking putative repeats within a genome, this pipeline

383   generates gene models, including 5′ and 3′ UTR's, by integrating *ab initio* gene predictions with aligned

384   transcript and protein evidence. First, we mapped and assembled transcripts using the RNA-Seq data from an

385   experiment of *N. vespilloides* in multiple behavioral states over a breeding cycle (mated, caring, and post-

386   caring; see Parker *et al.*, 2015 for full details) using the Bowtie (v2.2.3)/TopHat (v2.0.13)/Cufflinks (v2.2.1)

387   pipeline (Langmead *et al.*, 2010; Trapnell *et al.*, 2010; Kim *et al.*, 2013). To begin the annotation process, we

388   annotated the genome exclusively with the *N. vespilloides* Cufflinks-assembled transcripts and the proteomes

389   from five insects (*T. castaneum, N. vitripennis, A. mellifera, M. domestica, D. melanogaster*; downloaded

390   from UniProtKB, including all isoforms for comprehensive coverage) using default parameters, except for

391   est2genome=1, protein2genome=1. After this first iteration of annotation (and every subsequent iteration),

392   three scaffolds were inspected to visually check for annotation biases (Hoff and Stanke, 2015) using Apollo

12

393   genome browser (Lewis *et al.*, 2002). The next iteration used the same input data and parameters, except

394   changes to split_hit=2000, correct_est_fusion=1, which corrected for the smaller intron size observed and the

395   propensity of MAKER to fuse gene models that likely should be separate as inferred by visual inspection of

396   BLAST evidence. For the next iteration, three *ab initio* gene predictors were included in the annotation

397   process; Augustus (v2.5.5; Stanke et a., 2006), GeneMark-ES (v4.21; Lomsadze *et al.*, 2005), SNAP (v2010-

398   7-28; Korf, 2004; using est2genome=0, protein2genome=0). With AUGUSTUS, we used the "tribolium"

399   gene set provided with its distribution to guide gene predictions. GeneMark was trained on the draft

400   assembly of the *N. vespilloides* genome sequence using its automated training routine. SNAP was trained

401   using the MAKER2 gene models produced during the first round of annotation. All gene predictors were run

402   with default parameter values. The annotation was iterated twice with the gene predictors, updating the

403   SNAP HMM's between the two iterations. tRNA's were identified using tRNAscan-SE (v1.23; Lowe and

404   Eddy, 1997) within the Maker2 pipeline during the last iteration. Other non-coding RNA (rRNA, miRNA,

405   snRNA, snoRNA) were predicted and annotated with INFERNAL (v1.1.1; Nawrocki and Eddy, 2013) using

406   the complete Rfam database (v12.0; Daub *et al.*, 2014; Supplementary Table S4).

407

408   **Functional Annotation of Predicted Protein-Coding Genes.** To gain insight into the putative function of

409   each gene model, we annotated our gene models with three pipelines. First, we used BLASTp (v2.2.26;

410   Altschul *et al.*, 1997) to find the best hit against the entire UniProtKB database (vJan15; E-value: 10e-5).

411   Next, we used InterProScan (v5.8-49.0; Hunter *et al.*, 2009) to find the known protein domains within every

412   gene model from the TIGRFAM, ProDom, SMART, TMHMM, Phobius, PANTHER, PrositeProfiles,

413   SignalP-EUK, SuperFamily, PRINTS, Gene3d, PIRSF, Pfam, and Coils databases. We also used

414   InterProScan5 to assign Gene Ontology (GO) terms to further characterize each protein. KEGG pathway

415   analysis was also performed using the KEGG Automatic Annotation Sever (KAAS; Moriya *et al.*, 2007)

416   using the single-directional best hit method to assign orthology with default parameters and the default

417   Eukaryote gene sets plus all available arthropod gene sets.

418

419   **Ortholog Comparison.** To compare the orthology of our gene models to other insects, we analyzed our final

420   MAKER2 proteome using OrthoMCL (v2.0.9; Li *et al.*, 2003) against five other insect proteomes (*Tribolium*

421   *castaneum, Nasonia vitripennis, Apis mellifera, Musca domestica, Drosophila melanogaster*). If a gene was

422   represented by more than one isoform in its respective OGS, the longest isoform was chosen for this analysis.

423   We used BLASTp (E-value: 1e-5) to characterize the homology amongst all proteins. The output from this

424   analysis was used by OrthoMCL to cluster proteins into orthologous groupings. Results are presented as

425   Venn diagrams generated using the University of Ghent Bioinformatics Evolutionary Genomics' Venn

426   Diagram webtool (http://bioinformatics.psb.ugent.be/webtools/Venn/).

13

427

428   **Gene Family Expansion/Contraction Analysis.** To investigate possible expansion and contraction of

429   shared gene families of the six insects that we used in the OrthoMCL analysis, we used CAFÉ (v3.1; default

430   settings; Han *et al.*, 2013) with phylogenetic relationships from Trautwein *et al.* (2012) and divergence times

431   from TimeTree (Hedges *et al.*, 2006). Only gene families with at least one representative from *N.*

432   *vespilloides* were considered as gene family contractions.

433   Enrichment of gene ontology (GO) terms among the expanded gene family members was performed

434   using argiGO's web-based Singular Enrichment Analysis (Du *et al.*, 2010) of customized annotations by

435   comparing the GO terms associated with methylated gene from the InterProScan results to all GO terms

436   associated with all genes from InterProScan. Specifically, a hypergeometric test with a Benjamini-Hochberg

437   false discovery rate correction at a family-wise error rate of 0.05 was applied after GO terms were converted

438   into generic GO slim terms. All other parameters were set at default values.

439

440   **Selection Analysis.** To assess the rates of molecular evolution within the *N. vespilloides* genome, we used

441   PAML (Yang & Bielawski, 2000; Yang, 2007) to calculate *dN, dS,* and their ratio (*w*) and compare these

442   metrics to the *T. castaneum* and *D. ponderosae* lineages. We identified a set of 1-1 orthologs between *N.*

443   *vespilloides, D. ponderosae* and *T. castaneum* using a combination of the BLAST (Altschul *et al.*, 1990;

444   Camacho *et al.*, 2009), orthAgogue (Ekseth *et al.*, 2014), and mcl (Enright *et al.*, 2002; van Dongen, 2008) as

445   well as part of the OrthoMCL (Li *et al.*, 2003) pipeline. In total, 5,584 orthologs between all three species

446   were recovered. Amino acid sequences for each were aligned in PRANK (v100802; Löytynoja & Goldman,

447   2005). Codeml in the PAML package was used to test different models of molecular evolution for each gene.

448   Our interest is in determining which genes show evidence of a differential rate of evolution within *N.*

449   *vespilloides*. We therefore tested a basic model (model = 0, NSsites = 0, fix_omega = 0) that assumes a

450   single *w* across all the entire phylogeny against a branch model (model = 2, NSsites = 0, fix_omega = 0),

451   which assumes one *w* for the *N. vespilloides* branch and another *w* for the branches to *T. castaneum* and *D.*

452   *ponderosae*. These models are compared, for each gene, with a likelihood ratio test with one degree of

453   freedom. We then adjusted the significance threshold for a gene to show statistically significant different

454   rates of sequence evolution using a Benjamini-Hochberg false discovery rate (FDR) correction at *q* of 0.05

455   (Benjamini & Hochberg, 1995). Finally, any estimates of *dS, dN* or *w* > 10 were discarded. These species are

456   phylogenetically distant (240 Mya) and this increases the likelihood signals of molecular evolution will be

457   lost due to saturation of *dS*.

458

459   **DNA Methylation Analysis.** As the first step to characterize if DNA methylation existed within *N.*

460   *vespilloides*, we use BLASTp (Altschul *et al.*, 1997) to identify putative DNA methyltransferases. We search

14

461   our genome with known members of Dnmt families of both vertebrate (*Mus musculus*; 1, 2, 3a, 3b, 3l) and

462   invertebrate (*T. castaneum, A. mellifera, D. melanogaster*; 1, 2, and 3). After three putative loci were found

463   (one member per Dnmt family), we further characterized the possible functional relationship of the proteins

464   by clustering them with the BLAST query proteins and several more invertebrate species (*Zootermopsis*

465   *nevadensis* and *Camponotus flordanus*) using ClustalW followed by a Neighbor Joining tree with 10,000

466   bootstraps in CLC Sequence Viewer (v7.5; http://www.clcbio.com) with default settings.

467   To address if DNA methylation is present in *N. vespilloides* at all, we performed methylC-Seq

468   (Lister *et al.*, 2008), whole genome sequencing on bisulfite treated genomic DNA, on three biological

469   replicates of whole larvae to create single base resolution of DNA methylation, if present. DNA was

470   extracted from three whole *N. vespilloides* larvae respectively using the same protocol as for the Illumina and

471   PacBio sequencing (see above). Due to previous reports that *T. castaneum* contains no DNA cytosine

472   methylation (Zemach *et al.* 2010), samples from this species were used as negative control and DNA was

473   extracted from three biological replicates that each contained at least 15 pooled whole larvae using the same

474   protocol as for *N. vespilloides*. MethylC-Seq libraries were prepared according to the following the protocol

475   of Urich *et al.* (2015). Deep sequencing was performed using an Illumina NextSeq500 Instrument at the

476   University of Georgia Genomics Facility (Supplementary Table S5).

477   Raw fastq files were trimmed for adapters CutAdapt (v1.3) and preprocessed to remove low quality

478   reads. We aligned quality controlled reads to the *N. vespilloides* v1 and *T. castaneum* v3.0 reference

479   genomes, respectively, using the same method as previously described in (Schmitz *et al.*, 2013). The *T.*

480   *castaneum* genome and OGS gff (v3) were obtained from BeetleBase.org. Lambda sequence (which is fully

481   unmethylated) was used as control to calculate the efficiency of the sodium bisulfite reaction and the

482   associated non-conversion rate of unmodified cytosines, which ranged from 0.10-0.11% (Supplementary

483   Table S5). Only cytosine sites with a minimum coverage of at least three reads were used for subsequent

484   analysis. A binomial test coupled with Benjamini-Hochberg correction was adopted to determine the

485   methylation status of each cytosine. Weighted methylation levels were calculated as previously described

486   (Schultz *et al.*, 2012).

487   *Nicrophorus vespilloides* replicate 1 was used to compute the exact values and percentages and plot

488   Figure 5C, 5E, 5F, 5H, and 5G, but all replicates were qualitatively similar (Supplementary Table S7).

489   Methylated cytosines and their flanking two bases were extracted out for sequence conservation analysis

490   using the program WebLogo 3.3 (Crooks *et al.*, 2004). To perform symmetry analysis, both strands of each

491   CpG dinucleotide were required to have a minimum coverage of at least three reads and at least one of the

492   CpG sites was identified as methylated. Upstream regions were defined as 1kb upstream starting from the

493   translational start site or the transcriptional if a 5´ UTR was annotated. The program bedtools was used to

494   determine the distribution of methylated CpG sites (Quinlan and Hall, 2010). To identify "methylated"

15

495   versus "unmethylated" genes, a binomial test was performed on genes that contained at least five methylated

496   CpG sites. The weighted methylation level (total methylated cytosines / total mapped cytosines within a

497   region of interest) was determined for all genes that had at least five CpG sites and this was used as the

498   probability of a gene being methylated. These results were then corrected for multiple testing using

499   Benjamini-Hochberg correction with an FDR cut-off of 5% (Benjamini and Hochberg, 1995). For figure 5G,

500   each gene element was divided evenly into 20 bins and the weighted methylation levels were calculated for

501   each bin.

502   Enrichment of gene ontology (GO) terms among the methylated genes was tested as above.

503   All analyses were conducted on the zcluster housed at the Georgia Advanced Computing Resource

504   Center at the University of Georgia.

505

518

519   **References**

520   Altschul SF, Madden TL, Schäffer AA, Zhang J, Zhang Z, Miller W, Lipman DJ. 1997. Gapped BLAST and

521   PSI-BLAST: a new generation of protein database search programs. *Nuc Acids Res* 25: 3389–3402.

522

523   Bao Z, Eddy SR. 2002. Automated *de novo* identification of repeat sequence families in sequenced genomes.

524   *Genome Res* 12: 1269-1276.

525

526   Boake CRB, Arnold SJ, Breden F, Meffert LM, Ritchie MG, Taylor BJ, Wolf JB, Moore AJ. 2002. Genetic

527   tools for studying adaptation and the evolution of behavior. *Amer Nat* 160: S143-S159.

528

529   Benjamini Y, Hochberg Y. 1995. Controlling the false discovery rate: a practical and powerful approach to
530   multiple testing. *J Royal Stat Soc B (Meth.)* 57: 289-300.
531

532   Benson G. 1999. Tandem repeats finder: a program to analyze DNA sequences. *Nuc Acids Res* 27: 573.
533

534   Bonasio R. Li Q, Lian J, Mutti NS, Jin L, Zhao H, Zhang P, Wen P, Xiang H, Ding Y, Jin Z, et al. 2012.
535   Genome-wide and caste-specific DNA methylomes of the ants *Camponotus floridanus* and *Harpegnathos*
536   *saltator*. *Curr Biol* 22: 1755-1764.
537

538   Camacho C, Coulouris G, Avagyan V, Ma N, Papadopoulos, Bealer K, Madden TL. 2009. BLAST+:
539   architecture and applications. *BMC Bioinfo* 10, 421.
540

541   Campbell MS, Holt C, Moore B, Yandell, M. 2014. Genome annotation and curation using MAKER and
542   MAKER-P. *Curr Prot Bioinfo* 48: 4.11.1-4.11.39.
543

544   Cardoso SD, Teles MC, Oliveira RF. 2015. Neurogenomic mechanisms of social plasticity. *J Exp Biol* 218:
545   140-149.
546

547   Costa JT. 2006. *The Other Insect Societies*. (Harvard University Press, 2006).
548

549   Crooks GE, Hon G, Chandonia, JM, Brenner SE. 2004. WebLogo: a sequence logo generator. *Genome Res*
550   14: 1188-1190.
551

552   Cunningham CB, Douthit MK, Moore AJ. 2014. Octopaminergic gene expression and flexible social
553   behavior in the subsocial burying beetle *Nicrophorus vespilloides*. *Insect Mol Biol* 23: 391-404.
554

555   Dönitz J, Schmitt-Engel C, Grossman D, Gerischer L, Tech M, Schoppmeier M, Klingler M, Bucher G.
556   2015. iBeetle-Base: a database for RNAi phenotypes in the red flour beetle *Tribolium castaneum*. *Nuc*
557   *Acids Res* 43: D721-D726.
558

559   Dos Santos G. Schroeder AJ, Goodman JL, Strelets VB, Crosby MA, Thurmond J, Emmert DB, Gelbart
560   WM, FlyBase Consortium 2015. FlyBase: Introduction of the *Drosophila melanogaster* release 6
561   reference genome assembly and large-scale migration of genome annotations. *Nuc Acids Res* 43: D690-
562   D697.

563

564   Du Z, Zhou X, Ling Y, Zhang Z, Su Z. 2012. agriGO: a GO analysis toolkit for the agricultural community.
565   *Nuc Acids Res* 38: W64-W70.

566

567   Eggert A-K, Müller JK. 1997. Biparental care and social evolution in burying beetles: lessons from the
568   larder. In Choe JC, Crespi BJ, editors. *The Evolution of Social Behavior in Insects and Arachnids*.
569   Cambridge: Cambridge University Press. p216–236

570

571   Eggert A-K, Reinking M, Müller JK. 1998. Parental care improves offspring survival and growth in burying
572   beetles. *Anim Behav* 55: 97-107.

573

574   Ekseth OK, Kuiper M, Mironov V. 2014. OrthAgogue: An agile tool for the rapid prediction of orthology
575   relations. *Bioinfo* 30: 734–736. doi: 10.1093/bioinformatics/btt582.

576   Elsik CG, Worley KC, Bennett AK, Beye M, Camara F, Childers CP, Graaf DC, Debyser G, Deng J,
577   Devreese B, et al. 2014. Finding the missing honey bee genes: lessons learned from a genome upgrade.
578   *BMC Genomics* 15: 86.

579

580   English AC, Richards S, Han Y, Wang M, Vee V, Qu J, Qin X, Muzny DM, Reid JG, Worley KC, Gibbs RA
581   2012. Mind the gap: Upgrading genomes with Pacific Biosciences RS long-read sequencing technology.
582   *PLoS ONE* 7: e47768.

583

584   Enright AJ, Dongen SV, Ouzounis CA. 2002. An efficient algorithm for large-scale detection of protein
585   families. *Nuc Acids Res* 30: 1575–1584.

586

587   Falckenhayn C, Boerjan B, Raddatz G, Frohme M, Schoofs L, Lyko. 2013.Characterization of genome
588   methylation patterns in the desert locust *Schistocerca gregaria*. *J Exp Biol* 216: 1423-1429.

589

590   Fischman BJ, Woodard SH, Robinson GE. 2011. Molecular evolutionary analysis of insect societies. *Proc
591   Nat Acad Sci USA* 108: 10847-10854.

592

593   Gadau J, Helmkampf M, Nygaard S, Roux J, Simola DF, Smith CR, Suen G, Wurm Y, Smith CD. 2012. The
594   genomic impact of 100 million years of social evolution in seven ant species. *Trends Genetics* 28: 14-21.

595

18

596 Glastad KM, Chau LM, Goodisman MAD. 2015. Epigenetics in social insects. In Zayed A, Kent CF, editors.

597 *Advances in Insect Physiology*. Amsterdam: Elsevier. p227-269.

598

599 Han MV, Thomas GWC, Lugo-Martinez J, Hahn MW. 2013. Estimating gene gain and loss rates in the

600 presence of error in genome assembly and annotation using CAFE 3. *Mol Biol Evol* 30: 1987-1997.

601

602 Hare EE, Johnston JS. 2011. Genome size determination using flow cytometry of propidium iodide-stained

603 nuclei. In Orgogozo V, Rockman M, editors. *Molecular Methods for Evolutionary Genetics*. Humana

604 Press. p3-12.

605

606 Hedges SB, Dudley J, Kumar S. 2006. TimeTree: a public knowledge-base of divergence times among

607 organisms. *Bioinfo* 22: 2971-2972.

608

609 Herb BR, Wolschin F, Hansen KD, Aryee MJ, Langmead B, Irizarry R, Amdam GV, Feinberg AP. 2012.

610 Reversible switching between epigenetic states in honeybee behavioral subcastes. *Nature Neurosci* 10:

611 1371-1373.

612

613 Hoff KJ, Stanke M. 2015. Current methods for automated annotation of protein-coding genes. *Curr Op*

614 *Insect Sci* 7: 8-14.

615

616 Holt C, Yandell M. 2011. MAKER2: an annotation pipeline and genome-database management tool for

617 second-generation genome projects. *BMC Bioinfo* 12: 491.

618

619 Hopwood PE, Moore AJ, Royle NJ. 2013. Nutrition during sexual maturation affects competitive ability but

620 not reproductive productivity in burying beetles. *Func. Ecol.* 27: 1350-1357.

621

622 Hunt T, Bergsten J, Levkanicova Z, Papadopoulou A, John OS, Wild R,Hammond PM, Ahrens D, Balke M,

623 Caterino MS, Gomez-Zurita J, Ribera I, Barraclough TG, Bocakova M, Bocak L, Vogler. 2007. A

624 comprehensive phylogeny of beetles reveals the evolutionary origins of a supperradiation. *Science* 318:

625 1913-1916.

626

627 Hunter S, Apweiler R, Attwood TK, Bairoch A, Bateman A, Binns D, Bork P, Das U, Daugherty L,

628 Duquenne L, et al. 2009. InterPro: the integrative protein signature database. *Nuc Acids Res* 37: D211-

629 D215.

630

631 Jiang, N. 2014. ProtExcluder1.1.

632   http://weatherby.genetics.utah.edu/MAKER/wiki/index.php/Repeat_Library_Construction--Basic.

633

634 Kajitani R, Toshimoto K, Noguchi H, Toyoda A, Ogura Y, Okuno M, Yabana M, Harada M, Nagayasu E,

635   Maruyama H, Kohara Y, Fujiyama A, Hayashi T, Itoh T. 2014. Efficient de novo assembly of highly

636   heterozygous genomes from whole-genome shotgun short reads. *Genome Res* 24: 1384-1395.

637

638 Kapheim KM, Pan H, Li C, Salzberg SL, Puiu D, Magoc T, Robertson HM, Hudson ME, Venkat A,

639   Fischman BJ, et al. 2015. Genomic signatures of evolutionary transitions from solitary to group living.

640   *Science* 348: 1139-1143.

641

642 Keeling CI, Yuen MM, Liao NY, Docking TR, Chan SK, Taylor GA, Palmquist DL, Jackman SD, Nguyen

643   A, Li M, Henderson H, Janes JK, Zhao Y, Pandoh P, Moore R, Sperling FA, Huber DP, Birol I, Jones SJ,

644   Bohlmann J. 2013. Draft genome of the mountain pine beetle, *Dendroctonus ponderosae* Hopkins, a

645   major forest pest. *Genome Biol* 14: R27.

646

647 Kim D, Pertea G, Trapnell C, Pimentel H, Kelley R, Salzberg SL 2013. TopHat2: accurate alignment of

648   transcriptomes in the presence of insertions, deletions and gene fusions. *Genome Biol.* 14: R36.

649

650 Kim S, Murphy T, Xia J, Caragea D, Park Y, Beeman RW, Lorenzen MD, Butcher S, Manak JR, Brown SJ.

651   2010. BeetleBase in 2010: revisions to provide comprehensive genomic information for *Tribolium*

652   *castaneum*. *Nuc Acids Res* 38: D437–D442.

653

654 Kocher SD, Li C, Yang W, Tan H, Yi SV, Yang Z, Hoekstra HE, Zhang G, Pierce NE, Yu DW. 2013. The

655   draft genome of a socially polymorphic halictid bee, *Lasioglossum albipes*. *Genome Biol* 14: R142.

656

657 Koren S, Schatz MC, Walenz BP, Martin J, Howard JT, Ganapathy G, Wang Z, Rasko DA, McCombie WR,

658   Jarvis ED, Phillippy AM. 2012. Hybrid error correction and de novo assembly of single-molecule

659   sequencing reads. *Nature Biotech* 30: 693-700.

660

661 Korf I. 2004. Gene finding in novel genomes. *BMC Bioinfo* 5: 59.

662

663   Langmead B, Trapnell C, Pop M, Salzberg SL. 2009. Ultrafast and memory-efficient alignment of short
664   DNA sequences to the human genome. *Genome Biol* 10: R25.
665

666   Lewis SE, Searle SM, Harris N, Gibson M, Lyer V, Richter J, Wiel C, Bayraktaroglu L, Birney E, Crosby
667   MA, Kaminker JS, Matthews BB, Prochnik SE, Smithy CD, Tupy JL, Rubin GM, Misra S, Mungall CJ,
668   Clamp ME. 2002. Apollo: a sequence annotation editor. *Genome Biol* 3: research0082.1–0082.14.
669

670   Li L, Stoeckert Jr CJ, Roos DS. 2003. OrthoMCL: Identification of ortholog groups for eukaryotic genomes.
671   *Genome Res* 13: 2178-2189.
672

673   Lister R, O'Malley RC, Tonti-Filippini J, Gregpry BD, Berry CC, Millar H, Ecker JR. 2007. Highly
674   integrated single-base resolution maps of the epigenome in *Arabidopsis*. *Cell* 133: 523-36.
675

676   Lomsadze A, Ter-Hovhannisyan V, Chernoff YO, Borodovdky M. 2005. Gene identification in novel
677   eukaryotic genomes by self-training algorithm. *Nuc Acids Res* 33: 6494-6506.
678

679   Lowe TM, Eddy SR. 1997. tRNAscan-SE: A program for improved detection of transfer RNA genes in
680   genomic sequence. *Nucl Acids Res* 25: 955-964.
681

682   Löytynoja A, Goldman N. 2005. An algorithm for progressive multiple alignment of sequences with
683   insertions. *Proc Nat Acad Sci* 102: 10557–10562.
684

685   Lyko F, Foret S, Kucharski R, Wolf S, Falckenhayn C, Maleszka R. 2010. The honey bee epigenomes:
686   Differential methylation of brain DNA in queens and workers. *PLoS Biol* 9:
687   10.1371/journal.pbio.1000506.
688

689   Magoc T, Salzberg, S. 2011. FLASH: fast length adjustment of short reads to improve genome assemblies.
690   *Bioinfo* 27: 2957-2963.
691

692   Martin M. 2011. Cutadapt removes adapter sequences from high-throughput sequencing reads. *EMBnet* 17:
693   10-12.
694

695   McGlothlin JW, Moore AJ, Wolf JB, Brodie III ED. 2012. Interacting phenotypes and the evolutionary
696   process. III. Social evolution. *Evolution* 64: 2558-2574.

697

698 Mikheyev AS, Linksvayer TA. 2015. Genes associated with ant social behavior show distinct transcriptional
699   and evolutionary patterns. eLife 4: e04775.

700

701 Nawrocki EP, Eddy SR. 2013. Infernal 1.1: 100-fold faster RNA homology searches. *Bioinfo* 29: 2933-2935.

702

703 Moriya Y, Itoh M, Okuda S, Yoshizawa A, Kanehisa M. 2007. KAAS: An automatic genome annotation and
704   pathway reconstruction server. *Nuc Acids Res* 35: W182-W185.

705

706 Nawrocki EP, Burge SW, Bateman A, Daub J, Eberhardt RY, Eddy SR, Floden EW, Gardner PP, Jones TA,
707   Tate J, Finn RD. 2014. Rfam 12.0: Updates to the RNA families database. *Nuc Acids Res*
708   10.1093/nar/gku1063.

709

710 Oxley PR, Ji L, Fetter-Pruneda I, McKenzie SK, Li C, Hu H, Zhang G, Kronauer DJC. 2014. The genome of
711   the clonal raider ant *Cerapachys biroi*. *Curr Biol* 24: 451-458.

712

713 Parker D, Cunningham CB, Walling CA, Stamper CE, Head ML, Roy-Zokan E, McKinney EC, Ritchie MG,
714   Moore AJ. 2015. Gene expression versus phenotype: transcriptomes of parents help identify parenting
715   strategies and sexual conflict. *in review*.

716

717 Parra G, Bradnam K, Ning Z, Keane T, Korf I. 2009. Assessing the gene space in draft genomes. *Nuc Acids*
718   *Res* 37: 289-297.

719

720 Price AL, Jones NC, Pevzner PA. 2005. *De novo* identification of repeat families in large genomes. *Bioinfo*
721   21: i351-i358.

722

723 Quinlan AR, Hall IM. 2010. BEDTools: A flexible suite of utilities for comparing genomic features. *Bioinfo*
724   26: 841-842.

725

726 Rehan SM, Toth AL. 2015. Climbing the social ladder: the molecular evolution of sociality. *Trends Ecol*
727   *Evol* 30: 426-433.

728

729 Richards, S. 2015. It's more than just stamp collecting: how genome sequencing can unify biological
730   research. *Trends Genet* 34: 411-421.

731

732   Richards S, Murali SC. 2015. Best practices in insect genome sequencing: what works and what doesn't.

733   *Curr Op Insect Sci* 7: 1-7.

734

735   Roy-Zokan E, Cunningham CB, Hebb LE, McKinney EC, Moore AJ. 2015. Vitellogenin and vitellogenin

736   receptor gene expression is associated with male and female parenting in a subsocial insect. *Proc Royal*

737   *Soc B* 282: 20150787. doi: 10.1098/rspb.2015.0787.

738

739   Sadd BM, Barribeau SM, Bloch G, de Graaf DC, Dearden P, Elsik CG, Gadau J, Grimmelikhuijzen CJ,

740   Hasselmann M, Lozier JD, et al. 2015. The genomes of two key bumblebee species with primitive

741   eusocial organization. *Genome Biol* 16: 76.

742

743   Schmieder R, Edwards R. 2011a. Quality control and preprocessing of metagenomic datasets. *Bioinfo* 27:

744   863-864.

745

746   Schmieder R, Edwards R. 2011b. Fast identification and removal of sequence contamination from genomic

747   and metagenomic datasets. *PLoS ONE* 6: e17288.

748

749   Schmitz RJ, He Y, Valdes-Lopez O, Khan SM, Joshi T, Urich MA, Nery JR, Diers B, Xu D, Stacey G, Ecker

750   JR. 2013. Epigenome-wide inheritance of cytosine methylation variants in a recombinant inbred

751   population. *Genome Res* 23: 1663-1674.

752

753   Schultz MD, Schmitz RJ, Ecker JR. 2012. 'Leveling' the playing field for analyses of single-base resolution

754   DNA methylomes. *Trends Genetics* 28: 583-585.

755

756   Scott, MP. 1998. The ecology and behavior of burying beetles. *Ann Rev Entomol* 43: 595-618.

757

758   Shelton JM, Coleman MC, Herndon N, Lu N, Lam ET, Anantharaman, T, Sheth P, Brown SJ. 2015. Tools

759   and pipelines for BioNano data: molecule assembly pipeline and FASTA super scaffolding tool. *bioRxiv*.

760   doi: http://dx.doi.org/10.1101/020966.

761

762   Sikes DS, Venables C. 2013. Molecular phylogeny of the burying beetles (Coleoptera: Silphidae:

763   Nicrophorinae). *Mol Phylo Evol* 69: 552-565.

764

765   Smiseth PT, Darwell CT, Moore AJ. 2003. Partial begging: an empirical model for the early evolution of
766   offspring signaling. *Proc Royal Soc B* 270: 1773-1777.
767
768   Smiseth PT, Dawson C, Varley E, Moore AJ. 2005.How do caring parents respond to mate loss? Differential
769   response by males and females. *Anim Behav* 69: 551-559.
770
771   Smiseth PT, Lennox L, Moore AJ. 2007. Interaction between parental care and sibling competition: parents
772   enhance offspring growth and exacerbate sibling competition. *Evolution* 61: 2331-2339.
773
774   Smit AFA, Hubley R, Green P. 2015. RepeatMasker Open-4.0. 2013-2015. http://www.repeatmasker.org.
775
776   Smith CR, Smith CD, Robertson HM, Helmkampf M, Zimin A, Yandell M, Holt C, Hu H, Abouheif E, et al.
777   2012. Draft genome of the red harvester ant *Pogonomyrex barbatus*. *Proc Nat Acad Sci USA* 108: 5667-
778   5672.
779
780   Stanke M, Tzvetkova A, Morgenstern B. 2006. AUGUSTUS at EGASP: using EST, protein and genomic
781   alignments for improved gene prediction in the human genome. *Genome Biol* 7: S11.
782
783   Terrapon N, Li C, Robertson HM, Ji L, Meng X, Booth W, Chen Z, Childers CP, Glastad KM, Gokhale K, et
784   al. 2014. Molecular traces of alternative social organization in a termite genome. *Nature Comm* 5: 3636.
785
786   Trapnell C, Williams BA, Pertea G, Mortazavi A, Kwan G, van Baren MJ, Salzberg SL, Wold BJ, Pachter L.
787   2010. Transcript assembly and quantification by RNA-Seq reveals unannotated transcripts and isoform
788   switching during cell differentiation. *Nature Biotech* 28: 511–515.
789
790   Trautwein MD, Weigmann BM, Beutel R, Kjer KM, Yeates DK. 2012. Advances in insect phylogeny at the
791   dawn of the postgenomic era. *Ann Rev Entomol* 57: 449-468.
792
793   Tribolium Genome Sequencing Consortium. 2008. The genome of the model beetle and pest *Tribolium
794   castaneum*. *Nature* 452: 949-955.
795
796   Trumbo ST. 2012. Patterns of parental care in invertebrates. In Royle, NJ, Smiseth PT, Kölliker, editors. *The
797   Evolution of Parental Care* . Oxford: Oxford University Press. p81-100.
798

24

799   Trumbo ST, Rauter CM. 2014. Juvenile hormone, metabolic rate, body mass and longevity costs in parenting
800   burying beetles. *Anim Behav* 92: 203-211.
801

802   Urich MA, Nery JR, Lister R, Schmitz RJ, Ecker JR. 2015. MethylC-seq library preparation for base-
803   resolution whole-genome bisulfite sequencing. *Nature Prot* 10: 475-483.
804

805   van Dongen, S. 2008. Graph Clustering via a Discrete Uncoupling Process. *J Matrix Anal App* 30: 121–141.
806

807   Walling CA, Stamper CE, Smiseth PT, Moore AJ. 2008. The quantitative genetics of sex differences in
808   parenting. *Proc Nat Acad Sci USA* 105: 18430-18435.
809

810   Wang S, Lorenzen MD, Beeman RW, Brown SJ. 2008. Analysis of repetitive DNA distribution patterns in
811   the *Tribolium castaneum* genome. *Genome Biol* 9: R61.
812

813   Wang X, Wheeler D, Avery A, Rago A, Choi J-H, Colbourne JK, Clark AG, Werren JH. 2013. Function and
814   evolution of DNA methylation in *Nasonia vitripennis*. *PLoS Genet* 9: e1003872.
815

816   Wang X, Fang X, Yang P, Jiang X, Jiang F, Zhao D, Li B, Cui F, Wei J, Ma C, et al. 2014. The locust
817   genome provides insight into swarm formation an long-istance flight. *Nature Comm* 5: 2957.
818

819   Waterhouse RM, Zdobnov EM, Tegenfeldt F, Li J, Kriventseva EV. 2013. OrthoDB: a hierarchical catalog
820   of animal, fungal and bacterial orthologs. *Nuc Acids Res* 41: D358-65.
821

822   Wiegmann BM, Trautwein MD, Kim J-W, Cassel BK, Bertone MA, Winterton SL, Yeates DK. 2009.
823   Single-copy nuclear genes resolve the phylogeny of the holometabolous insects. *BMC Biol* 7: 34.
824

825   Wilson EO. 1971. *The Insect Societies*. Cambridge: Harvard University Press.
826

827   Wurm Y, Wang J, Riba-Grognuz O, Corona M, Nygaard S, Hunt BG, Ingram KK, Falquet L,
828   Nipitwattanaphon M, Gotzek D, et al. 2012. The genome of the first ant *Solenopsis invicta*. *Proc Nat*
829   *Acad Sci USA* 108: 5679-5684.
830

831   Xiang H, *et al.* 2010. Single base-resolution methylome of the silkworm reveals a sparse epigenomic map.
832   *Nature Biotech* 28: 516-520.

833

834   Yan H, Simola DF, Bonasio R, Leibig J, Berger SL, Reinberg D. 2014. Eusocial insects as emerging models

835   for behavioral epigenetics. *Nature Rev Genet* 15: 677-688.

836

837   Yang Z. 2007. PAML 4: phylogenetic analysis by maximum likelihood. *Mol Biol Evol* 24: 1586–1591.

838

839   Yang Z, Bielawski JP. 2000. Statistical methods for detecting molecular adaptation. *Trends Ecol Evol* 15:

840   496–503.

841

842   Yu P, McKinney EC, Kandasamy MM, Albert AL, Meagher RB. 2015. Characterization of brain cell nuclei

843   with decondensed chromatin. *Dev Neurobiol* 75: 738-756.

844

845   Zemach A, McDaniel IE, Silva P, Zilberman D. 2010. Genome-wide evolutionary analysis of eukaryotic
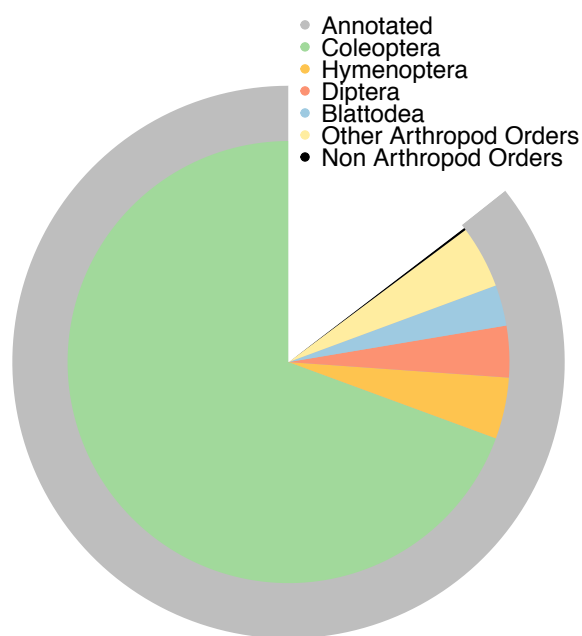
846   DNA methylation. *Science* 328: 916-919.

847

848   **Figures**

849

850   Figure 1.

851

852

853   Fig. 1. An adult female *Nicrophorus vespilloides* regurgitating food into the mouth of her begging larvae on a

854   prepared mouse carcass. Photograph by A. J. Moore.

855   Figure 2.

856

857   Fig. 2. A two-ring pie chart showing results of annotation with BLAST against the complete UniProtKB

858   database. First outer ring (grey) shows the proportion of gene models that could be annotated. Second ring

859   (multi-colored) shows the proportion of best BLAST hits of the annotations by Order for all species with 5 or

860   more best hits (97.8%). The best BLAST hits were overwhelmingly from other beetles and other Arthropods.

861

862   Figure 3.
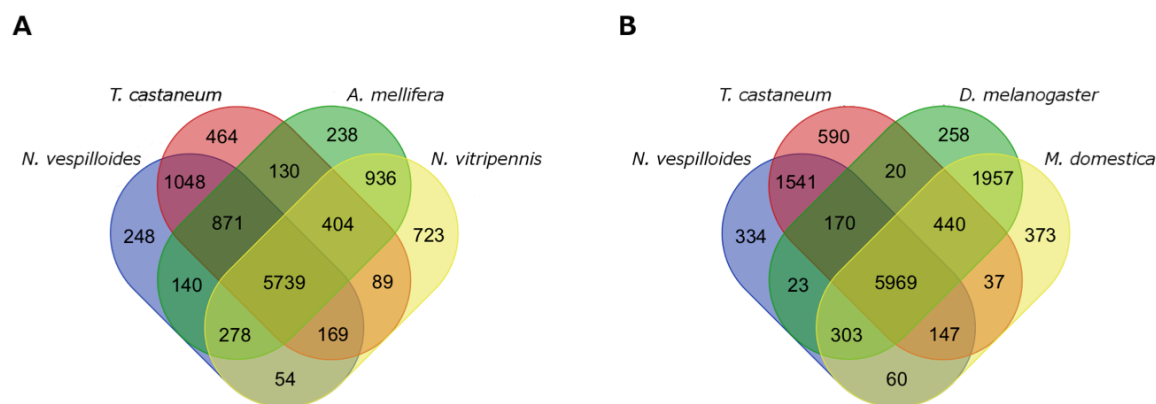
863

864   Fig. 3. Figure shows the results of the OrthoMCL analysis that clustered the proteomes of *Nicrophorus*

865   *vespilloides*, *Tribolium castanuem*, *Apis mellifera*, *Nasonia nitripennis*, *Drosophila melanogaster*, and

27

866 *Musca domestica* into orthologous groupings. (A) A Venn diagram showing the overlap in the orthologous

867 groupings of the two beetles (*T. castaneum* and *N. vespilloides*) and the two Hymenoptera (*A. mellifera* and

868 *N. vitripennis*). (B) A Venn diagram showing the overlap in orthologous groupings of the two beetles (*T.*

869 *castaneum* and *N. vespilloides*) and the two Diptera (*D. melanogaster* and *M. domestica*).

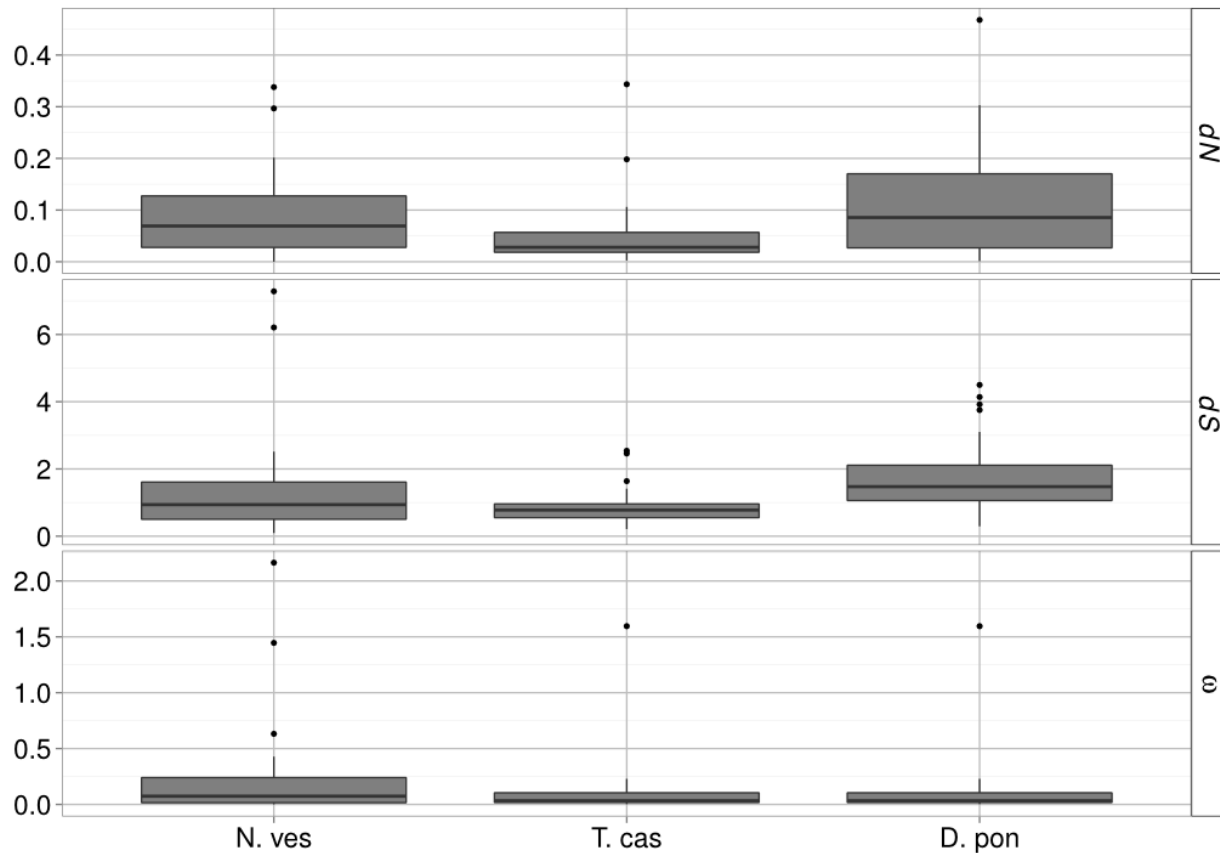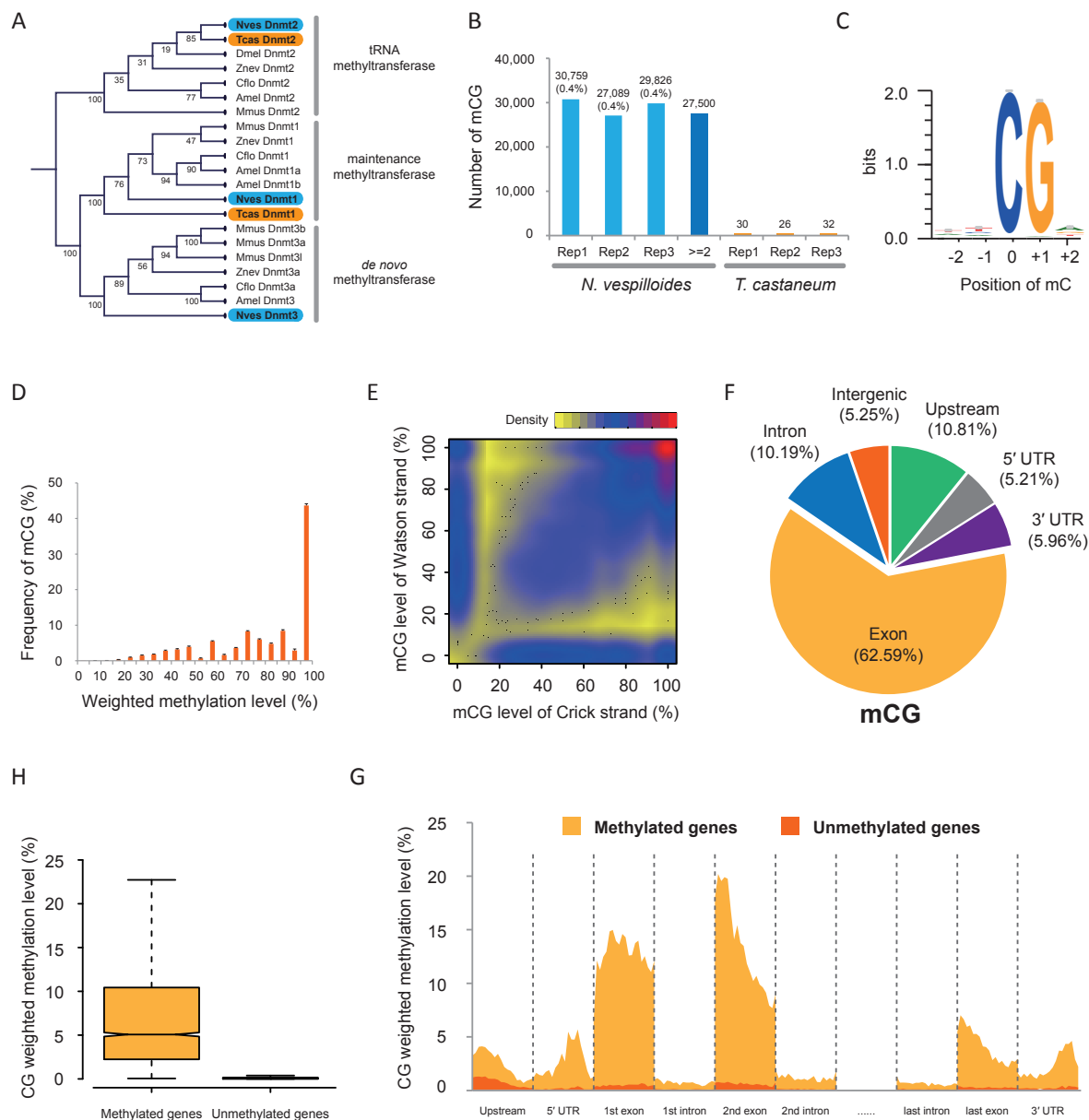870

871 Figure 4

872

873 Fig. 4. Distributions of *dN, dS* and *ω* for 25 genes, filtered for *dN, dS* or *ω* < 10, that show evidence of

874 differential rates of sequence evolution on the *Nicrophorus vespilloides* (N. ves) lineage in comparison to the

875 *Tribolium castaneum* (T. cas) and *Dendroctonus ponderosae* (D. pon) lineages.

28

876   Figure 5

877

878

879   Fig. 5. Summary of DNA methylation analyses. (A) A cladogram showing the relationship of the Dnmt's

880   across several insects and a mammal. Nves = *Nicrophorus vespilloides*, Tcas = *Tribolium castaneum*, Dmel =

881   *Drosophila melanogaster*, Cflo = *Camponotus floridanus*, Amel = *Apis mellifera*, Mmus = *Mus musculus*.

882   (B) Number of methylated cytosines in each of the three replicates of *N. vespilloides* and *T. castaneum*. (C)

883   A sequence logo of the overwhelming occurrence of methylation in CpG dinucleotide by showing the

884   nucleotide proportions of the two nucleotides both upstream and downstream of the methylated cytosines.

885   (D) A histogram of the percentage of statistically methylated CpG's versus their weighted methylation level

886   showing the typical pattern of a cytosine either being methylated or non-methylated. (E) A density plot

887   showing the very high symmetry of methylated CpG sites on opposing strands of DNA. (F) A pie chart

888   showing the distribution of methylated CpG's across gene elements. (H) A standard box plot of the

889   methylation level of genes grouped by whether a gene was significantly more methylated compared to the

890   average gene showing the typical bimodal where genes are either methylated or non-methylated. (G)

891   Diagram showing the weighted methylation level of each region of a gene model summarized as 20 bins

892   within a region. Genes were separated into "methylated" and "unmethylated" genes as in H and plotted

893   separately.

894

895   Tables

896

Table 1. Summary Statistics of N. vespilloides Draft Genome Assembly.

Total Assembled Length (bp)	195,308,655
Contigs (n)	5,858
Contig N50 (bp)	102,139
Largest Contig (bp)	944,646
Scaffolds (n)	4,664
Scaffold N50 (bp)	122,407
Largest Scaffold (bp)	1,795,199
% GC Content	31.85
Predicted Gene Models (n of loci)	13,526
CEGMA Pipeline Analysis (%- complete/partial)	99.6/100
T. castaneum BUSCO (%- identity/Recovered/Single Copy)	65.95/90.0/96.8
BioNano Genomics Genome Map Alignment to *in silico* maps (%)	31

897
898