



OPEN

## Exploring bycatch diversity of organisms in whole genome sequencing of Erebidae moths (Lepidoptera)

Hamid Reza Ghanavi<sup>1</sup>✉, Victoria G. Twort<sup>1,2</sup> & Anne Duplouy<sup>1,3</sup>

Models estimate that up to 80% of all butterfly and moth species host vertically transmitted endosymbiotic microorganisms, which can affect the host fitness, metabolism, reproduction, population dynamics, and genetic diversity, among others. The supporting empirical data are however currently highly biased towards the generally more colourful butterflies, and include less information about moths. Additionally, studies of symbiotic partners of Lepidoptera predominantly focus on the common bacterium *Wolbachia pipientis*, while infections by other inherited microbial partners have more rarely been investigated. Here, we mine the whole genome sequence data of 47 species of Erebidae moths, with the aims to both inform on the diversity of symbionts potentially associated with this Lepidoptera group, and discuss the potential of metagenomic approaches to inform on host associated microbiome diversity. Based on the result of Kraken2 and MetaPhlan2 analyses, we found clear evidence of the presence of *Wolbachia* in four species. Our result also suggests the presence of three other bacterial symbionts (*Burkholderia* spp., *Sodalis* spp. and *Arsenophonus* spp.) in three other moth species. Additionally, we recovered genomic material from bracovirus in about half of our samples. The detection of the latter, usually found in mutualistic association to braconid parasitoid wasps, may inform on host-parasite interactions that take place in the natural habitat of the Erebidae moths, suggesting either contamination with material from species of the host community network, or horizontal transfer of members of the microbiome between interacting species.

A growing scientific community now sees each organism as a community of interacting species rather than as an independent entity. Insects are no exception. They host a variety of microbial symbionts sitting both inside and outside their host cells. These microorganisms are at least as numerous as the number of host cells, and may constitute up to 10% of the host total mass<sup>1</sup>. The effects of symbionts on their insect hosts are potentially as diverse as their taxonomy, ranging from pathogenic to obligate mutualists, and all the intermediate possible relationships<sup>2</sup>. This diversity has recently attracted the growing interest of the scientific community, but gaps and biases remain. For example, in Lepidoptera, research in symbiosis has mostly focused on the most charismatic groups of colourful diurnal butterflies<sup>3–5</sup> and on pest species to the human society<sup>6–8</sup>. In contrast, the rest of Lepidoptera (mainly moths), which encompass no less than 130,000 species<sup>9</sup> (80% of all Lepidoptera), have rarely been screened for their associations with symbionts<sup>10</sup>.

High throughput sequencing technologies (HTS) now provide a relatively easy and cheap way to obtain large amounts of genetic data. These technologies used to generate genomic data are varied and broadly applicable to the widest range of organisms. Thereby, revolutionizing our accessibility to genomic resources and continually expanding and renewing the scope of the questions we can address within the natural sciences. For example, sequencing material from a particular study organism, either entirely or partially, may result in a mix of primary host specific DNA and DNA from other sources. These other sources can include ecto/endosymbionts, food, opportunistic parasites and pathogens, among others. Such genomic data opens up the genomic analyses towards broader targets, including towards investigating the diversity of symbionts that might be associated to particular targeted hosts.

<sup>1</sup>Department of Biology, Lund University, Ecology Building, Sölvegatan 37, 22362 Lund, Skåne, Sweden. <sup>2</sup>The Finnish Museum of Natural History, Luomus, Zoology Unit, The University of Helsinki, Helsinki, Finland. <sup>3</sup>Insect Symbiosis Ecology and Evolution, Organismal and Evolutionary Biology Research Program, The University of Helsinki, Helsinki, Finland. ✉email: hamid.ghanavi@biol.lu.se

Here, we mine the data produced from whole genome sequencing of 47 moth species from the hyper-diverse family Erebidae (24,000 species) to (1) explore the potential diversity of symbionts associated to this megadiverse Lepidoptera family; and (2) to evaluate the exploratory power of recovering information on natural host-symbiont associations from the low coverage genome sequencing approaches.

## Results

**Metagenomic analysis.** We identified the species *Idia aemula*, *Luceria striata*, *Acantholipes circumdata* and *Oraesia excavata* (RZ271, RZ42, RZ248, and RZ337) as infected by *Wolbachia*, and *Wolbachia*-associated phage WO (Table 1), with between 66,978 and 208,044 of the reads identified as belonging to the symbiont. Additionally, the reads obtained from sample RZ13 (*Gonitis involuta*) was also found to include 954 *Wolbachia* reads, which is a higher number of reads than found for any of the clearly uninfected specimens, but is considerably less than any of the four clearly infected specimens listed above. The mapping of the reads to two known *Wolbachia* reference genomes (*wMel*, GCF\_000008025.1 and *wPip*, GCF\_000073005.1) show a relatively homogeneous coverage of the reference genomes (Fig. 1) with mean coverages between 10 and 40 times the reference genome (Table 2). In the case of RZ13 sample, even though the coverage seemed homogeneously scattered through the reference genome, the mean coverage was lower than 1x (Table 2).

Both Kraken2 and MetaPhlan2 analyses showed no to very few reads mapping to *Cardinium*, *Hamiltonella* or *Spiroplasma* bacteria, or to Microsporidian fungi, in any of the 47 datasets screened. In contrast, the specimens RZ103 and RZ111 (*Rema costimacula* and *Platyjionia mediorufa*) included considerably more reads from *Sodalis* bacteria (9108 and 4395, respectively), and from *Arsenophonus* bacteria (1336 and 662, respectively), than any other samples (maximum of 50 reads in any other sample). A closer look at the Kraken2 outputs from the latter two samples also revealed a possible infection with a *Plautia stali* symbiont (gammmaproteobacteria; 3856 and 1914 reads, respectively), which was not detected in any of the other 45 samples. Additionally, the sample RZ30 (*Cretonotos transiens*) is the only one to show relatively high number of reads mapping to *Burkholderia* bacteria (N = 1995). Finally, we identified a considerable amount of reads from viruses of the *Polydnaviridae* family, and especially of the bracoviruses in three samples, *Erebus ephesperis*, *Masca abactalis* and *Asota heliconia* (RZ11, 1288 reads, RZ18, 1381 reads, and RZ44, 1384 reads). All other samples only include less than 750 reads, and more often no reads, for these viruses.

All details of the screen for the common symbionts can be found in Table 1, while all results from the Kraken2 and MetaPhlan2 analyses can be found in the supplementary material and GitHub repository.

## Discussion

We confidently add four moth species (i.e., *Idia aemula*, *Luceria striata*, *Acantholipes circumdata* and *Oraesia excavata*) to the list of species hosting the intracellular alpha-proteobacterial symbiont *Wolbachia*<sup>10</sup>, confirmed through two screening methods (i.e., Kraken2 and MetaPhlan2). With only 4 out of 47 species (8%) found infected, this represents a lower infection rate than the current literature suggests (i.e., 16–79% of the studied lepidopteran groups infected with *Wolbachia*<sup>11–16</sup>). The general penetrance of *Wolbachia* however varies significantly among species, and is often low within infected populations<sup>17</sup>. Thus, with only one sample screened per species, our results are most likely underestimating the true infection rate within the Erebidae moths. Future broader screenings of different populations will provide more accurate natural infection rates for these species. Although microbial surveys in *Calyptra thalictra*<sup>18</sup> and *Lymantria dispar*<sup>19,20</sup> did not highlight *Wolbachia* infections in these species, a recent screening of diverse moth species from Thailand, showed that two (22%; *Olepa* sp. and *Cretonotos transiens*) out of nine Erebidae species screened (ie. *Amata* sp., *Asota plana*, *Cretonotos transiens*, *Euplocia memblaria*, *Fodina contigua*, *Neochera inops*, *N. dominia*, *Olepa* sp., *Pareuchaetes pseudoinvoluta*) were infected by the bacterial symbiont<sup>21</sup>.

Noticeably, we observe the presence of *Wolbachia* phage WO within the samples for which *Wolbachia* presence is strongly supported. The interaction of this bacteriophage with *Wolbachia* has been the focus of many evolutionary studies in recent years<sup>22–26</sup>. Previous research suggests that phage WO are associated with horizontal gene transfer in *Wolbachia*, and with genes that may affect the fitness of the bacterium<sup>27,28</sup>. These bacteriophages have been observed in practically all the studied genomes of *Wolbachia* up to date, with very few obligate mutualistic exceptions<sup>22,29,30</sup>. In the sample RZ13, species *Gonitis involuta*, a relatively high number of reads mapped to *Wolbachia* (1 K reads), although significantly lower than in the other four species (29–144 K reads), and no reads were mapped to phage-WO. In addition to the relatively lower sequencing depth compared to the other positive cases, few non-excluding hypotheses may explain such a pattern, these reads might originate from (1) contamination with other genetic material alien to our sample, (2) the integration of *Wolbachia* genomic material (partially or entirely) in the host genome, (3) random errors in the identification of the reads as *Wolbachia*, (4) low quality genomic material or (5) a combination of above-mentioned reasons. The overall screening results suggest that this sample was of low quality prior to sequencing. We however cannot rule out any of the other possibilities, and more studies are needed to fully confirm or reject the presence of *Wolbachia* in this species.

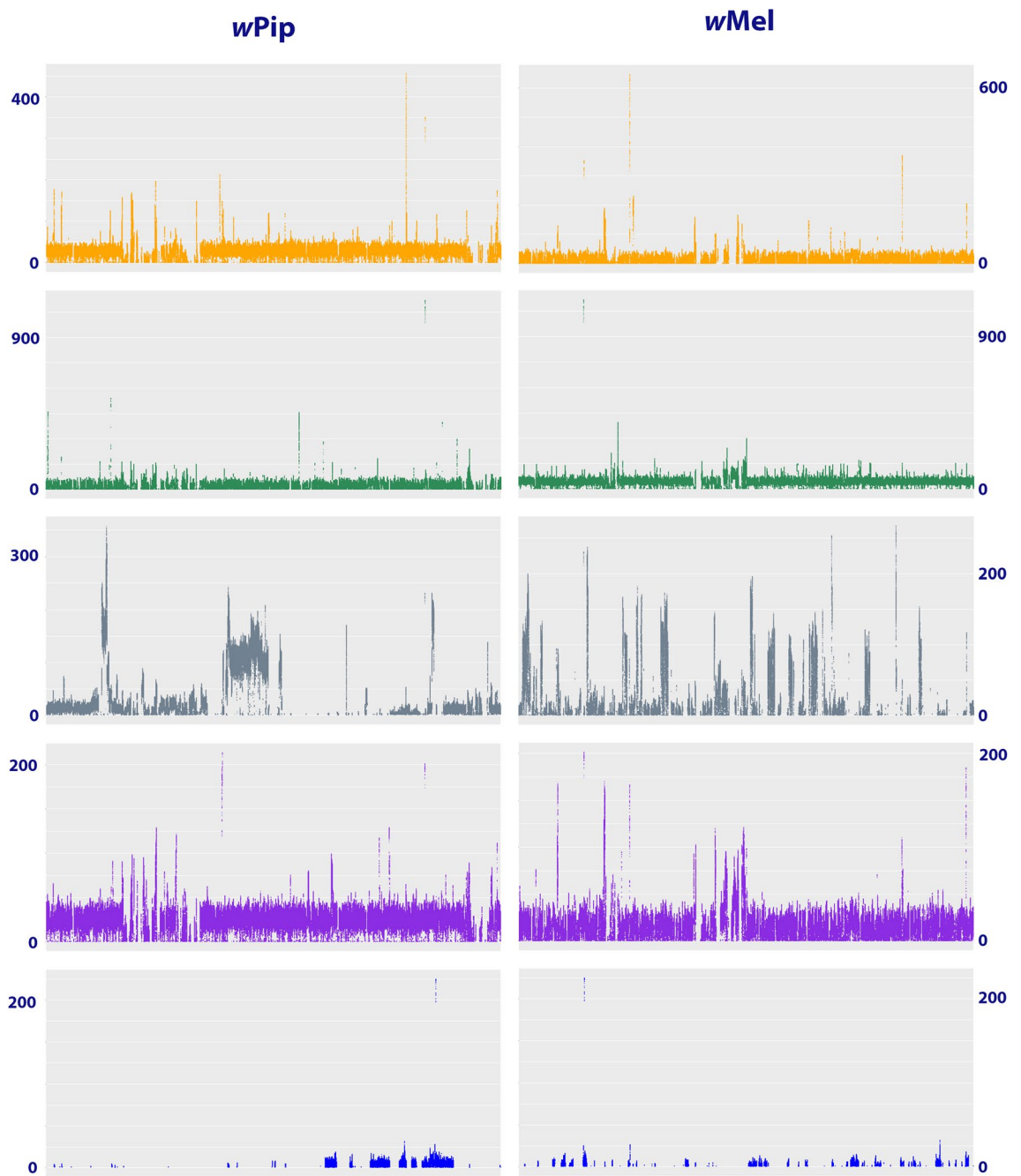
The two samples, *Rema costimacula* (RZ103) and *Platyjionia mediorufa* (RZ111), were of particular interests. Both the Kraken2 and the MetaPhlan2 analyses suggest the presence of three gammaproteobacteria endosymbionts, namely *Sodalis*, *Arsenophonus* and ‘*Plautia stali*-symbiont’ in both samples. *Sodalis* has been characterized from different insects, including tsetse flies<sup>31</sup>, seal louse<sup>32</sup>, pigeon louse<sup>33</sup>, loose flies<sup>34</sup>, aphids<sup>35</sup>, seed bug<sup>36</sup>, weevils<sup>37,38</sup>, stinkbugs<sup>39</sup>, bees<sup>40</sup>, and ants<sup>41</sup>, among others. To our best knowledge however, this is the first time the three symbionts are found in Lepidoptera (Duploux and Hornett<sup>10</sup>). This suggests that *Sodalis* bacteria might affect a more diverse group of organisms than is currently known. We are however cautious with the interpretation of this result, as the simple discovery of bacteria in the genomic data does not inform us about the nature of their interactions with the hosts. Whether *Sodalis* and the moth species share a symbiotic relationship, or

Ref	Code	Species	Country	# raw reads (Million)	Kraken2 Results											Metaphlan2 results	
					Lepidoptera	Spiroplasma	Burkholderia	Sodalis	Arsenophonus	Rickettsia	Wolbachia	Wolbachia PhageWo	Ichnovirus	Bracovirus	Microsporidia		Wolbachia
1	MM00407	<i>Scoliopteryx libatrix</i>	FINLAND	38	2,266,973	-	289	-	-	-	-	-	-	-	-	-	-
2	RZ103	<i>Rema costimacula</i>	HONG KONG	22	907,037	-	-	9,108	1,336	-	-	-	-	-	266	-	-
3	RZ104	<i>Saroba pustulifera</i>	HONG KONG	21	1,649,430	-	-	-	-	-	-	-	-	-	-	-	-
4	RZ105	<i>Hypocala deflorata</i>	HONG KONG	48	3,231,681	-	59	-	-	-	-	-	-	-	-	-	-
5	RZ11	<i>Erebus ephesperis</i>	TAIWAN	106	8,550,697	64	298	-	-	-	-	-	-	-	1,288	-	-
6	RZ111	<i>Platylonia mediorufa</i>	HONG KONG	26	995,385	-	-	4,395	662	-	-	-	-	-	-	-	-
7	RZ119	<i>Schistorhynchus argentipectus</i>	HONG KONG	56	5,928,236	56	99	-	-	-	-	-	-	-	-	-	-
8	RZ13	<i>Gonitis involuta</i>	TANZANIA	17	1,254,304	-	83	-	-	-	-	954	-	317	102	-	2,005
9	RZ138	<i>Micronoctua</i> sp.	INDONESIA	107	11,736,010	100	126	50	-	-	-	-	-	-	-	-	-
10	RZ149	<i>Hypopyra capensis</i>	GHANA	53	4,808,838	-	107	-	-	-	-	-	-	-	-	-	-
11	RZ159	<i>Rivula ochrea</i>	GHANA	59	6,499,556	71	216	-	-	-	-	-	-	-	-	-	-
12	RZ18	<i>Masca abactalis</i>	INDONESIA	45	4,175,988	-	67	-	-	-	-	-	-	-	1,381	-	-
13	RZ180	<i>Nodaria verticalis</i>	GHANA	38	4,198,076	-	116	-	-	-	-	-	-	-	1,731	-	-
14	RZ21	<i>Ophiura coronata</i>	MALAYSIA	42	2,653,381	-	76	-	-	-	-	-	-	-	-	-	-
15	RZ22	<i>Azeta ceramina</i>	COSTA RICA	55	4,926,573	64	85	-	-	-	-	-	-	-	-	-	-
16	RZ248	<i>Acantholipes circumdata</i>	UAE	28	3,085,527	-	-	-	-	-	-	29,454	410	-	-	-	220,309
17	RZ265	<i>Rhesala imparata</i>	HONG KONG	38	6,206,848	-	67	-	-	-	-	-	-	-	-	-	-
18	RZ268	<i>Meodina praecipua</i>	HONG KONG	26	2,200,296	-	-	-	-	-	-	-	-	-	790	-	-
19	RZ271	<i>Idia aemula</i>	USA	52	6,897,287	-	112	-	-	-	-	144,331	1,038	-	771	-	168,228
20	RZ28	<i>Brunia antica</i>	HONG KONG	77	7,118,395	59	242	-	-	50	-	-	-	-	-	-	-
21	RZ3	<i>Laspeyria flexula</i>	HUNGARY	54	7,583,217	-	82	-	-	-	-	161	-	-	-	-	-
22	RZ30	<i>Cretonotos transiens</i>	HONG KONG	30	6,196,702	-	1,995	-	-	-	-	-	-	-	198	-	-
23	RZ313	<i>Synoides fumosa</i>	JAPAN	87	10,986,269	-	505	-	-	-	-	-	-	-	104	-	576
24	RZ331	<i>Tinolius eburneigutta</i>	THAILAND	33	3,112,193	-	85	-	-	-	-	-	-	-	159	-	-
25	RZ332	<i>Anoba anguliplaga</i>	GHANA	42	1,874,468	-	79	-	-	-	-	-	-	-	-	-	-
26	RZ336	<i>Calyptra hokkaida</i>	JAPAN	34	5,835,726	-	122	-	-	-	-	-	-	-	341	-	-
27	RZ337	<i>Oraesia excavata</i>	HONG KONG	38	3,147,679	-	65	-	-	-	-	66,978	182	-	581	-	208,044
28	RZ34	<i>Nygmia plana</i>	HONG KONG	19	1,026,248	-	-	-	-	-	-	-	-	-	-	-	-
29	RZ367	<i>Hypera baltimoralis</i>	USA	35	3,005,435	-	54	-	-	-	-	-	-	-	-	-	-
30	RZ389	<i>Tamsia hieroglyphica</i>	MALAYSIA	26	1,285,828	-	63	-	-	-	-	-	-	-	572	-	-
31	RZ39	<i>Ericia subcinerea</i>	HONG KONG	80	7,549,078	-	133	-	-	-	-	65	-	-	-	-	-
32	RZ4	<i>Colobochyla salicalis</i>	HUNGARY	44	5,510,176	-	57	-	-	-	-	-	-	-	72	-	-
33	RZ40	<i>Pangrapta bicornuta</i>	HONG KONG	63	7,415,193	-	316	-	-	-	-	60	-	-	107	-	-
34	RZ404	<i>Amerila astruus</i>	MALAYSIA	45	4,649,942	-	72	-	-	-	-	-	-	-	130	-	-
35	RZ41	<i>Metaceme atrigutta</i>	HONG KONG	17	1,079,839	-	-	-	-	-	-	-	-	-	-	-	-
36	RZ42	<i>Luceria striata</i>	HONG KONG	27	3,065,608	-	75	-	-	-	-	67,176	494	-	-	-	181,728
37	RZ44	<i>Asota heliconia</i>	HONG KONG	40	3,763,381	-	54	-	-	-	-	-	-	-	1,384	-	-
38	RZ48	<i>Sympis rufibasis</i>	HONG KONG	52	5,491,409	-	208	-	-	-	-	-	-	-	-	-	-
39	RZ56	<i>Phylloides eyndhovii</i>	TAIWAN	64	4,058,586	-	118	-	-	-	-	-	-	-	-	-	-
40	RZ57	<i>Lygephila maxima</i>	JAPAN	41	3,832,732	-	117	-	-	-	-	-	-	-	-	-	-
41	RZ58	<i>Melipotis jucunda</i>	USA	57	5,764,266	-	101	-	-	-	-	-	-	-	101	-	-
42	RZ59	<i>Panopoda rufimargo</i>	USA	42	4,715,473	-	237	-	-	-	-	-	-	-	-	-	-
43	RZ8	<i>Syntomis phegea</i>	HUNGARY	22	1,675,147	-	-	-	-	-	-	-	-	-	309	177	-

Continued

Ref	Code	Species	Country	# raw reads (Million)	Kraken2 Results											Metaphlan2 results
					Lepidoptera	Spiroplasma	Burkholderia	Sodalis	Arsenophonus	Rickettsia	Wolbachia	Wolbachia PhageWo	Ichnovirus	Bracovirus	Microsporidia	
44	RZ89	Arctornis sp.	JAPAN	33	3,256,478	-	50	-	-	-	-	-	-	94	-	-
45	RZ9	Scolecocampa liburna	USA	52	3,132,323	-	137	-	-	-	-	-	-	-	-	-
46	RZ93	Epitausa dilina	COSTA RICA	41	4,202,446	-	-	-	-	-	-	-	-	-	-	-
47	RZ94	Alesua etialis	COSTA RICA	16	1,605,058	-	76	-	-	-	-	-	-	-	-	-

**Table 1.** The number of reads classified as originating from the host and various microorganisms. Values in bold highlight the values mentioned in the text, - represent samples with either zero or less than 50 reads classified.



**Figure 1.** The mapped reads to wMel and wPip Wolbachia reference genomes. The coverage is shown on the vertical side of the figure. The top graphs (yellow) correspond to the sample RZ337 (*Oraesia excavata*), followed by RZ271 (*Idia aemula* in green), RZ248 (*Acantholipes circumdata* in grey), RZ42 (*Luceria striata* in purple) and at the bottom RZ13 (*Gonitis involuta* in blue).

Sample	Reference	Reads mapped	Covered bases	%	Mean coverage
RZ13	Wmel	4102	129,338	10.2	0.4
	Wpip	8712	214,530	14.5	0.7
RZ248	Wmel	95,663	453,378	35.8	9.3
	Wpip	212,353	774,465	52.2	18.5
RZ271	Wmel	413,711	1,124,837	88.7	39.4
	Wpip	196,202	908,917	61.3	15.6
RZ337	Wmel	127,490	808,811	63.8	11.4
	Wpip	267,514	1,157,313	78.1	20.8
RZ42	Wmel	110,993	831,256	65.6	10.8
	Wpip	212,555	1,139,840	76.9	18.2

**Table 2.** Samples screened for *Wolbachia* genomes. The *wMel* and *wPip* are reference strains for the A- and B-*Wolbachia* supergroups, respectively. Covered bases column gives the length of the reference covered in bp. The column marked with % gives the percentage of the genome recovered.

not, will only be confirmed via experimentation and testing of the partnership through the host generations. Contamination of those two samples prior to DNA extraction is always possible. However, the sequenced host genetic material did not include significant amount of hemipteran DNA (or any other non-lepidopteran insect order), with comparable low numbers of reads (< 1500) mapped to hemipterans in all the sequenced genomes. This rules out DNA contamination by material from the previously confirmed hemipteran hosts of these three symbionts. It is shown that the female brown-winged green bug, *P. stali*, smears excrement over the egg surface during oviposition. The nymphs acquire the symbionts right after hatching by ingesting the excrements<sup>42</sup>. Therefore, a possible contamination source could be any contact with such excrement/egg clusters. Once again, studies of the symbionts in natural populations of these moth species are needed to fully resolve the true infection state of these species and the relationship with the bacteria.

The moth species *Cretonotos transiens* shows a potential partnership with proteobacteria *Burkholderia* sp. Recently, Boonsit and Wiwatanaratanabutr<sup>21</sup> found *Wolbachia* in 75% of the *C. transiens* samples they screened for (N = 6/8). Their samples were collected from Thailand, while the *C. transiens* specimen we analysed in this study originated from Hong Kong. In Lepidoptera, *Burkholderia* are known from the microbiota associated with the moth *Lymantria dispar*<sup>43</sup>. However, similarly to the other symbionts presented above, these bacteria are also found in very diverse groups of organisms, from Amoebas to Orthoptera, from humans to plants<sup>44–47</sup>. In the bean bug, *Riptortus pedestris*, studies have suggested that the bacteria can benefit their host by providing resistance to pesticides<sup>48</sup>. Although never tested, the presence of such Proteobacteria in moths could similarly enhance the host ability to resist pesticides. If proven true, this could contribute to partially explaining the global success of many pest moth species despite the development of various targeted control strategies.

Six genomes included significantly high amounts of bracovirus reads, *Erebus ephesperis* (RZ11), *Masca abacotalis* (RZ18), *Nodaria verticalis* (RZ180), *Mecodina praecipua* (RZ268), *Idia aemula* (RZ271) and *Asota heliconia* (RZ44). Bracoviruses are a known genus of mutualistic viruses with a complex life cycle. Integrated in the genome of a braconid parasitic wasp, the bracovirus is transcribed during oviposition in lepidopteran larvae<sup>49</sup>. The presence of this viral genetic material in adult moths might suggest an unsuccessful infection by the parasitoid, and the survival of the larvae carrying the parasitic viral particles. Another potential explanation includes the possibility for the viral DNA to be integrated into the lepidopteran genome, as it is usually found in its common Hymenoptera host. Only studies simultaneously investigating parasitism success rate and tissue tropism of the bracoviruses in the Lepidoptera and Hymenoptera hosts, will be able to inform on the nature of these interactions.

From a methodological point of view, the present study shows the successful exploratory approach to mine for potentially hidden associated microbial diversity in genomic data. Our study was performed on shallow genome short reads obtained using Illumina platform. The original purpose of this sequencing effort was to study the phylogenomics of the hosts species<sup>50</sup>, but a similar approach to the one we have taken here can be implemented to any publicly available genomic datasets. The popularity of genomic scale sequence data methods, such as Illumina short read approach, created a wide publicly open genomic resource for the research community to study questions that are not directly into the focus of the studies generating them. It is however important to also consider the limitations of such approaches. First, the quality and completeness of the reference datasets needed for programs like Kraken2 are bound to significantly affect the results. Second, incomplete and shallow genomes tend to present false negatives when mined for many symbionts. In addition, the origin of the DNA used for the genome sequencing will affect any conclusion on presence/absence or abundance of the symbionts detected and those undetected. In our study, all the used genomes came from DNA extracted from legs, therefore there is a methodical hard bias against gut fauna for example, however as shown in other studies some symbionts as *Wolbachia* are also found in the haemolymph of arthropods<sup>51</sup>. Third, this kind of exploratory analyses of genomic material does not inform about the nature of the interaction between the organisms found in the genomic mix. Furthermore, in the majority of cases, this method also does not inform on the origin of the organisms. This is especially important as sample contamination is a known problem since the appearance of the molecular sequencing techniques. Finally, this method is not suitable for quantification of the present organisms. Altogether, these limitations exemplify the exploratory nature of the approach we used in this study, and that we at best provide suspicion for diverse symbiotic infections in different Erebiidae moth species, which

presence and importance will only be fully confirmed via direct screening, and ecological and evolutionary studies of natural populations.

## Conclusion

As we expected, our method detects various symbiotic partners in several Erebidae moth species, including *Wolbachia* and the bacteriophage *WO* in four species, *Burkholderia* in one other species, and *Sodalis* and *Arsenophonus* simultaneously in two species. Although symbiotic associations of Lepidoptera with *Wolbachia* is likely, similar long-term associations between the three other symbionts and the Lepidoptera have yet to be described. Similarly, we detect DNA material from bracoviruses that are currently only described as mutualistic symbionts of Hymenoptera. The true nature of these associations requires further experimental investigation. The detection of bracovirus DNA could for example suggest ecological interactions between moths and parasitoids, and the ability of the formers to naturally resist parasitoid attack strategies. Altogether our study presents a method and produces material supporting testable hypotheses about the diversity and nature of symbiotic interactions in those particular Lepidoptera species. With the availability of open access metagenomics databases, this field promises extensive and exciting opportunities to explore potentially hidden symbiotic diversity.

## Material and methods

**Genome data.** We used the data produced from the whole genome sequencing project of 47 Erebidae species (see<sup>50</sup>). The sampling information is shown in Table 1. This selection includes genomes representing the main described subfamilies and major lineages within the Erebidae family. The DNA was extracted from one or two legs of the selected samples. Extractions took place in 2000 s/over a decade ago, for the purpose of another study (see<sup>52</sup>). It is important to keep in mind that the genome sequencing approach generating this dataset is not optimized to recover the symbiont diversity of these organisms, therefore the diversity is likely to be systematically underestimated.

**Metagenomic analysis.** The raw reads were quality checked with FASTQC v0.11.8<sup>53</sup>. Reads containing ambiguous bases were removed from the dataset using Prinseq 0.20.4<sup>54</sup>. Reads were cleaned to remove low quality bases from the beginning (LEADING: 3) and end (TRAILING: 3) and reads less than 30 bp in length. The evaluation of read quality with a sliding window approach was done in Trimmomatic 0.38<sup>55</sup>. Quality was measured for sliding windows of 4 bp and had to be greater than PHRED 25 on average. Cleaned reads were assigned taxonomic labels with Kraken2<sup>56</sup> and MetaPhlAn 2.0<sup>57</sup>. Kraken2 was run using a custom database, which contained the standard kraken database, the refseq viral, bacteria and plasmid databases and all available Lepidoptera genomes from genbank (Supplementary Table 1 contains a full list of taxa included), confidence threshold of 0.05, and a mpa style output. MetaPhlAn was run using the analysis type rel\_ab\_w\_read\_stats, which provides the relative abundance and an estimate of read numbers originating from each clade. We visually screened the result for each sample, focusing on seven genera of vertically transmitted bacterial symbionts (i.e., *Arsenophonus* sp., *Cardinium* sp., *Hamiltonella* sp., *Rickettsia* sp., *Sodalis* sp., *Spiroplasma* sp. and *Wolbachia* sp.), one group of fungal symbionts (Microsporidia), and three types of viral symbionts (i.e., *Wolbachia*-phage *WO*, ichnovirus and bracovirus). This represents a non-exhaustive list of the maternally inherited symbionts found in diverse insect hosts, but covers all of those that have already been characterized within Lepidoptera<sup>10</sup>. We also checked on the presence of the gut bacteria *Burkholderia* sp., which are known to confer pesticide resistance to their host in the pest bean bug *Riptortus pedestris* (e.g., ‘can degrade an organophosphate pesticide, fenitrothion’)<sup>58</sup>.

To discriminate between true and false positives a mapping analysis was carried out. For *Wolbachia* positive samples (list), cleaned reads were mapped to both the *wMel* (GCF\_000008025.1) and *wPip* (GCF\_000073005.1) genomes using bowtie2 v2.4.1<sup>59</sup> (sensitive local option). The resulting sam files were converted to sorted bam files with samtools v1.10<sup>60</sup>. Coverage information was obtained using samtools depth, and the resulting graphs plotted with ggplot package<sup>61</sup> in R.

## Data availability

The genome data used in this study are deposited in the NCBI SRA under BioProject PRJNA702831. All data in the supplementary material, the tables and the results can be found and downloaded from the GitHub repository: [github.com/Hamidhrg/ErebidSymbionts](https://github.com/Hamidhrg/ErebidSymbionts).

Received: 23 July 2021; Accepted: 30 November 2021

Published online: 30 December 2021

## References

- Douglas, A. E. Multiorganismal insects: Diversity and function of resident microorganisms. *Annu. Rev. Entomol.* **60**, 17–34 (2015).
- Dillon, R. J. & Dillon, V. M. THE gut bacteria of insects: Nonpathogenic interactions. *Annu. Rev. Entomol.* **49**, 71–92 (2004).
- Duplouy, A., Hurst, G. D. D., O’neill, S. L. & Charlat, S. Rapid spread of male-killing *Wolbachia* in the butterfly *Hypolimnys bolina*. *J. Evol. Biol.* **23**, 231–235 (2010).
- Altizer, S. M., Oberhauser, K. S. & Brower, L. P. Associations between host migration and the prevalence of a protozoan parasite in natural populations of adult monarch butterflies. *Ecol. Entomol.* **25**, 125–139 (2000).
- Jiggins, X., Hurst, X., Dolman, X. & Majerus, X. High-prevalence male-killing *Wolbachia* in the butterfly *Acraea encedana*. *J. Evol. Biol.* **13**, 495–501 (2000).
- Xu, P., Liu, Y., Graham, R. I., Wilson, K. & Wu, K. *Densovirus* is a mutualistic symbiont of a global crop pest (*Helicoverpa armigera*) and protects against a baculovirus and Bt Biopesticide. *PLoS Pathog.* **10**, e1004490 (2014).
- Bapatla, K. G., Singh, A., Yedula, S. & Patil, R. H. Annotation of gut bacterial taxonomic and functional diversity in *Spodoptera litura* and *Spilosoma obliqua*. *J. Basic Microbiol.* **58**, 217–226 (2018).

8. Chen, F. *et al.* Effects of Wolbachia on mitochondrial DNA variation in populations of *Aethis lepigone* (Lepidoptera: Noctuidae) in China. *Mitochondrial DNA Part A* **28**, 826–834 (2017).
9. van Nieuwerkerken, E. J. *et al.* Order Lepidoptera Linnaeus, 1758. In: Zhang, Z.-Q. (Ed.) Animal biodiversity: An outline of higher-level classification and survey of taxonomic richness. *Zootaxa* **1758**, 212–221 (2011).
10. Duploux, A. & Hornett, E. A. Uncovering the hidden players in Lepidoptera biology: The heritable microbial endosymbionts. *PeerJ* **6**, e4629 (2018).
11. Werren, J. H., Windsor, D. & Guo, L. Distribution of Wolbachia among neotropical arthropods. *Proc. R. Soc. Lond. Ser. B Biol. Sci.* **262**, 197–204 (1995).
12. Salunke, B. K. *et al.* Determination of Wolbachia diversity in butterflies from Western Ghats, India, by a multigene approach. *Appl. Environ. Microbiol.* **78**, 4458–4467 (2012).
13. Duploux, A. & Brattström, O. Wolbachia in the genus *Bicyclus*: A forgotten player. *Microb. Ecol.* **75**, 255–263 (2018).
14. Jiggins, F. M., Bentley, J. K., Majerus, M. E. & Hurst, G. D. How many species are infected with Wolbachia? Cryptic sex ratio distorters revealed to be common by intensive sampling. *Proc. R. Soc. Lond. Ser. B Biol. Sci.* **268**, 1123–1126 (2001).
15. Tagami, Y. & Miura, K. Distribution and prevalence of Wolbachia in Japanese populations of Lepidoptera. *Insect Mol. Biol.* **13**, 359–364 (2004).
16. Ilinsky, Y. & Kosterin, O. E. Molecular diversity of Wolbachia in Lepidoptera: Prevalent allelic content and high recombination of MLST genes. *Mol. Phylogenet. Evol.* **109**, 164–179 (2017).
17. Sazama, E. J., Ouellette, S. P. & Wesner, J. S. Bacterial endosymbionts are common among, but not necessarily within, insect species. *Environ. Entomol.* **48**, 127–133 (2019).
18. Zaspel, J. M. Systematics, biology, and behavior of fruit-piercing and blood-feeding moths in the subfamily calpinae (Lepidoptera: noctuidae). (2008).
19. Mason, C. J. & Raffa, K. F. Acquisition and structuring of midgut bacterial communities in gypsy moth (Lepidoptera: Erebidae) larvae. *Environ. Entomol.* **43**, 25 (2014).
20. Ilinsky, Y. *et al.* Detection of bacterial symbionts (Wolbachia, Spiroplasma) and eukaryotic pathogen (Microsporidia) in Japanese population of gypsy moth species (*Lymantria* spp.). *Euroasian Entomol. J.* **16**, 1–5 (2017).
21. Boonsit, P. & Wiwatanaratnabutr, I. Infection density, diversity, and distribution of Wolbachia bacteria in moths (Order Lepidoptera): First systematic report from Thailand. *J. Asia-Pac. Entomol.* **24**, 20 (2021).
22. Gavotte, L. *et al.* A survey of the bacteriophage WO in the endosymbiotic bacteria Wolbachia. *Mol. Biol. Evol.* **24**, 427–435 (2006).
23. Wang, G. H. *et al.* Bacteriophage WO can mediate horizontal gene transfer in endosymbiotic wolbachia genomes. *Front. Microbiol.* **7**, 1–16 (2016).
24. Wang, N., Jia, S., Xu, H., Liu, Y. & Huang, D. Multiple horizontal transfers of bacteriophage WO and host wolbachia in fig wasps in a closed community. *Front. Microbiol.* **7**, 1–10 (2016).
25. Tanaka, K., Furukawa, S., Nikoh, N., Sasaki, T. & Fukatsu, T. Complete WO phage sequences reveal their dynamic evolutionary trajectories and putative functional elements required for integration into the wolbachia genome. *Appl. Environ. Microbiol.* **75**, 5676–5686 (2009).
26. Kaushik, S., Sharma, K. K., Ramani, R. & Lakhanpaul, S. Detection of Wolbachia phage (WO) in Indian Lac insect [*Kerria lacca* (Kerr)] and its implications. *Indian J. Microbiol.* **59**, 237–240 (2019).
27. LePage, D. P. *et al.* Prophage WO genes recapitulate and enhance Wolbachia-induced cytoplasmic incompatibility. *Nature* **543**, 243–247 (2017).
28. Shropshire, J. D., On, J., Layton, E. M., Zhou, H. & Bordenstein, S. R. One prophage WO gene rescues cytoplasmic incompatibility in *Drosophila melanogaster*. *Proc. Natl. Acad. Sci.* **115**, 4987 (2018).
29. Bordenstein, S. R. & Bordenstein, S. R. Eukaryotic association module in phage WO genomes from Wolbachia. *Nat. Commun.* **7**, 13155 (2016).
30. Kent, B. N. & Bordenstein, S. R. Phage WO of Wolbachia: Lambda of the endosymbiont world. *Trends Microbiol.* **18**, 173–181 (2010).
31. Dale, C., Young, S. A., Haydon, D. T. & Welburn, S. C. The insect endosymbiont *Sodalis glossinidius* utilizes a type III secretion system for cell invasion. *Proc. Natl. Acad. Sci.* **98**, 1883–1888 (2001).
32. Boyd, B. M. *et al.* Two bacterial genera, *Sodalis* and *Rickettsia*, associated with the seal louse *Proechinophthirus fluctus* (Phthiraptera: Anoplura). *Appl. Environ. Microbiol.* **82**, 3185–3197 (2016).
33. Fukatsu, T. *et al.* Bacterial endosymbiont of the slender pigeon louse, *Columbicola columbae*, allied to endosymbionts of grain weevils and tsetse flies. *Appl. Environ. Microbiol.* **73**, 6660–6668 (2007).
34. Šochová, E., Husník, F., Nováková, E., Halajian, A. & Hýpša, V. *Arsenophonus* and *Sodalis* replacements shape evolution of symbiosis in louse flies. *PeerJ* **5**, e4099 (2017).
35. Burke, G. R., Normark, B. B., Favret, C. & Moran, N. A. Evolution and diversity of facultative symbionts from the aphid subfamily Lachninae. *Appl. Environ. Microbiol.* **75**, 5328–5335 (2009).
36. Santos-Garcia, D., Silva, F. J., Morin, S., Dettner, K. & Kuechler, S. M. The all-rounder *Sodalis*: A new bacteriome-associated endosymbiont of the lygaeoid bug *Henestaris halophilus* (Heteroptera: Henestarinae) and a critical examination of its evolution. *Genome Biol. Evol.* **9**, 2893–2910 (2017).
37. Toju, H. & Fukatsu, T. Diversity and infection prevalence of endosymbionts in natural populations of the chestnut weevil: Relevance of local climate and host plants. *Mol. Ecol.* **20**, 853–868 (2011).
38. Conord, C. *et al.* Long-term evolutionary stability of bacterial endosymbiosis in Curculionioidea: Additional evidence of symbiont replacement in the Dryophthoridae family. *Mol. Biol. Evol.* **25**, 859–868 (2008).
39. Kaiwa, N. *et al.* Bacterial symbionts of the giant jewel stinkbug *Eucorysses grandis* (Hemiptera: Scutelleridae). *Zool. Sci.* **28**, 169–174 (2011).
40. Rubin, B. E. R., Sanders, J. G., Turner, K. M., Pierce, N. E. & Kocher, S. D. Social behaviour in bees influences the abundance of *Sodalis* (Enterobacteriaceae) symbionts. *R. Soc. Open Sci.* **5**, 180369 (2018).
41. Sameshima, S., Hasegawa, E., Kitade, O., Minaka, N. & Matsumoto, T. Phylogenetic comparison of endosymbionts with their host ants based on molecular evidence. *Zool. Sci.* **16**, 993–1000 (1999).
42. Oishi, S., Moriyama, M., Koga, R. & Fukatsu, T. Morphogenesis and development of midgut symbiotic organ of the stinkbug *Plautia stali* (Hemiptera: Pentatomidae). *Zool. Lett.* **5**, 16 (2019).
43. Mason, C. J. & Raffa, K. F. Acquisition and structuring of midgut bacterial communities in gypsy moth (Lepidoptera: Erebidae) larvae. *Environ. Entomol.* **43**, 595–604 (2014).
44. Khojandi, N., Haselkorn, T. S., Eschbach, M. N., Naser, R. A. & DiSalvo, S. Intracellular Burkholderia symbionts induce extracellular secondary infections; driving diverse host outcomes that vary by genotype and environment. *ISME J.* **13**, 2068–2081 (2019).
45. Itoh, H. *et al.* Host–symbiont specificity determined by microbe–microbe competition in an insect gut. *Proc. Natl. Acad. Sci. USA* **116**, 22673–22682 (2019).
46. Ohbayashi, T., Itoh, H., Lachat, J., Kikuchi, Y. & Mergaert, P. Burkholderia gut symbionts associated with European and Japanese populations of the dock bug *Coreus marginatus* (Coreoidea: Coreidae). *Microbes Environ.* **34**, 219–222 (2019).
47. Itoh, H. *et al.* Evidence of environmental and vertical transmission of Burkholderia symbionts in the oriental chinch bug, *Cavelerius saccharivorus* (Heteroptera: Blissidae). *Appl. Environ. Microbiol.* **80**, 5974–5983 (2014).
48. Kikuchi, Y. *et al.* Symbiont-mediated insecticide resistance. *Proc. Natl. Acad. Sci. USA* **109**, 8618–8622 (2012).

49. Louis, F. *et al.* The bracovirus genome of the parasitoid wasp *Cotesia congregata* is amplified within 13 replication units, including sequences not packaged in the particles. *J. Virol.* **87**, 9649–9660 (2013).
50. Ghanavi, H. R., Twort, V., Hartman, T. J., Zahiri, R. & Wahlberg, N. The (non) accuracy of mitochondrial genomes for family level phylogenetics: The case of erebid moths (Lepidoptera; Erebidae). *bioRxiv* <https://doi.org/10.1101/2021.07.14.452330> (2021).
51. Rigaud, T. & Juchault, P. Success and failure of horizontal transfers of feminizing *Wolbachia* endosymbionts in woodlice. *J. Evol. Biol.* **8**, 25 (1995).
52. Zahiri, R. *et al.* Molecular phylogenetics of Erebidae (Lepidoptera, Noctuoidea). *Syst. Entomol.* **37**, 102–124 (2012).
53. Andrews, S. FastQC: A Quality Control Tool for High Throughput Sequence Data. <http://www.bioinformatics.babraham.ac.uk/projects/fastqc/> (2010).
54. Schmieder, R. & Edwards, R. Quality control and preprocessing of metagenomic datasets. *Bioinformatics* <https://doi.org/10.1093/bioinformatics/btr026> (2011).
55. Bolger, A. M., Lohse, M. & Usadel, B. Trimmomatic: A flexible trimmer for Illumina sequence data. *Bioinformatics* <https://doi.org/10.1093/bioinformatics/btu170> (2014).
56. Wood, D. E. & Salzberg, S. L. Kraken: Ultrafast metagenomic sequence classification using exact alignments. *Genome Biol.* **15**, R46 (2014).
57. Segata, N. *et al.* Metagenomic microbial community profiling using unique clade-specific marker genes. *Nat. Methods* **9**, 811–814 (2012).
58. Kikuchi, Y. & Yumoto, I. Efficient colonization of the bean bug *Riptortus pedestris* by an environmentally transmitted Burkholderia Symbiont. *Appl. Environ. Microbiol.* **79**, 2088–2091 (2013).
59. Langmead, B. & Salzberg, S. L. Fast gapped-read alignment with Bowtie 2. *Nat. Methods* **9**, 20 (2012).
60. Li, H. *et al.* The sequence alignment/map format and SAMtools. *Bioinformatics* **25**, 20 (2009).
61. Wickham, H. *ggplot2: Elegant Graphics for Data Analysis* (Springer, 2009).

## Acknowledgements

HRG received funding from the European Union's Horizon 2020 research and innovation program under the Marie Skłodowska-Curie Grant agreement No. 6422141. AD was funded by Marie Skłodowska-Curie Individual fellowship (#790531, Home Sweet Home). The authors acknowledge support from the National Genomics Infrastructure in Genomics Production Stockholm funded by Science for Life Laboratory, the Knut and Alice Wallenberg Foundation and the Swedish Research Council, and SNIC/Uppsala Multidisciplinary Center for Advanced Computational Science for assistance with massively parallel sequencing and access to the UPPMAX computational infrastructure. We acknowledge the international character of our team. It includes researchers from different countries, backgrounds, and career stages.

## Author contributions

H.R.G. and V.T. conceived of the presented idea. H.R.G. carried out the experiments and wrote the manuscript with input from all authors. V.T. and A.D. designed the computational framework and analysed the data. All authors discussed the results and commented on the manuscript.

## Funding

Open access funding provided by Lund University.

## Competing interests

The authors declare no competing interests.

## Additional information

**Supplementary Information** The online version contains supplementary material available at <https://doi.org/10.1038/s41598-021-03327-3>.

**Correspondence** and requests for materials should be addressed to H.R.G.

**Reprints and permissions information** is available at [www.nature.com/reprints](http://www.nature.com/reprints).

**Publisher's note** Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.



**Open Access** This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

© The Author(s) 2021