Repeated horizontal acquisition of lagriamide-producing symbionts in Lagriinae beetles

Short title: Selective beetle symbiont acquisition

Siddharth Uppal*[1], Samantha C. Waterworth*[1,2], Alina Nick (anick@ice.mpg.de)[3], Heiko Vogel (hvogel@ice.mpg.de)[3], Laura V. Flórez[4], Martin Kaltenpoth[3], Jason C. Kwan[1†]

*These authors contributed equally to this work

[†]To whom correspondence should be addressed. Address: 777 Highland Ave., Madison, WI 53705. Email: jason.kwan@wisc.edu

[1]Division of Pharmaceutical Sciences, School of Pharmacy, University of Wisconsin-Madison, Madison, USA.

[2]Current address: National Cancer Institute, Frederick, Maryland, USA

[3]Max Planck Institute for Chemical Ecology, Jena, Germany

[4]Department of Plant and Environmental Science, University of Copenhagen, Copenhagen, Denmark

Study funding information: NSF DBI-1845890 (S.U., S.C.W., J.C.K.)

Abstract

Microbial symbionts associate with multicellular organisms on a continuum from facultative associations to mutual codependency. In the oldest symbioses there is exclusive vertical symbiont transmission, and extensive co-diversification of symbiotic partners over millions of years and speciation events. Such symbionts often undergo genome reduction due to low effective populations, frequent population bottlenecks and weakened purifying selection. We describe here a group of Lagriinae beetle symbionts, which produce defensive lagriamide molecules to protect beetle eggs and larvae from fungal infection. These beetles likely evolved specialised symbiont-storage structures between 55 and 82 million years ago. Previous work found, despite the lagriamide-producing bacterium not being genetically isolated and potentially surviving in the environment, it underwent genome reduction in *Lagria villosa* hosts. Here, we use shotgun metagenomics to assemble 11 additional lagriamide-producing symbionts from diverse hosts within Lagriinae from five countries, to clarify the evolution of this symbiotic relationship. Surprisingly, we did not find evidence of co-evolution between symbionts and hosts, and although the lagriamide gene cluster (*lga*) had been horizontally transferred to the *L. villosa* symbiont, it likely was not transferred between symbionts. Instead, *lga* was likely obtained only once in the common ancestor of all lagriamide-producing symbionts and subsequently lost in related free-living strains. Our results suggest that beetles acquired lagriamide-producing symbionts specifically multiple times independently, which was followed by genome erosion. We also show that these symbionts have been replaced in the beetle host multiple times. The system therefore offers a unique opportunity to study the process of genome reduction.

Keywords: lagriamide, *Burkholderia*, symbiosis, symbiont replacement, biosynthetic gene cluster, metagenomics, Lagriinae, chemical defence, secondary metabolism

Introduction

Eukaryotic organisms have been associated with prokaryotic microbes since the initial endosymbiotic events which led to the acquisition of mitochondria and chloroplasts [1]. These organelles represent the presumed endpoint of ancient symbioses with α-proteobacteria and cyanobacteria, respectively, that over time led to a progressive shrinkage of the symbiont's genome and eventual transfer of genes from symbiont to host [1]. Although organelle acquisition appears to be a rare event [2], other more recent symbioses appear to be on a similar evolutionary trajectory of profound genome reduction and absolute dependence on host cells. For example, the acquisition of the intracellular symbiont *Buchnera aphidicola* in the common ancestor of aphids allowed them to diversify as sap-feeding insects since the symbiont synthesises several essential amino acids not found in plant sap, as evidenced by a rapid basal radiation of aphid species [3] and strict co-evolution of aphids and *Buchnera* [4]. *B. aphidicola* has been vertically transmitted for at least 200 million years [4] and has a profoundly reduced chromosome, about 11% of the size of that of *Escherichia coli* [5]. Through comparison of various symbionts, a model of genome reduction has emerged whereby host-restriction initially weakens purifying selection on formerly essential genes, through both host-provided metabolites and symbiont population structure, with low effective populations isolated in individual hosts [6]. When symbionts are vertically transmitted, population bottlenecks occur every time a host procreates, causing the fixation of deleterious mutations within the population [6]. These factors combine to first cause a proliferation of non-functional pseudogenes in the genome [6] and then deletion of those pseudogenes due to a known deletion-bias within bacteria [7]. The most reduced genomes lose even central functions such as DNA repair pathways [6], which leads to an increased rate of evolution and further gene loss, as well as increased AT-bias in many cases [8, 9]. In the cases of symbionts living inside host cells, it is easy to suppose that this process is exacerbated due to a lack of opportunity or ability to obtain

functional genes through homologous recombination - i.e. a combination of genetic isolation and disruption of homologous recombination apparatus.

However, genome reduction is also known to occur in the absence of genetic isolation. For instance, free-living bacteria living in nutrient poor environments such as *Prochlorococcus* spp. are thought to have reduced genomes as a consequence of selection pressure to streamline their metabolism [10], potentially explained through the Black Queen hypothesis [11], which posits that selection drives pathways to be lost when they are produced by another species in the ecosystem as “public goods”. There are also genome-reduced symbionts which seemingly are not genetically isolated. *Burkholderia* symbionts that reside extracellularly in leaf nodules in plants are mainly transmitted vertically because the symbiosis is mutually co-dependent [12], although horizontal transfer may have occurred rarely between plants, the soil microbiota and insects [13]. This suggests a lack of genetic isolation, and indeed there is evidence of repeated horizontal transfers of biosynthetic genes for defensive molecules among leaf nodule symbionts of *Rubiacaeae* plants [13]. Such systems may provide an opportunity to study the evolutionary pressures that lead to the process of genome reduction, and the mechanisms of symbiosis that underlie it.

The dichotomy of vertical versus horizontal transfer of symbionts may be one determinant of genome reduction. A relatively clear-cut example are two symbionts of the tunicate *Lissoclinum patella*, the extracellular cyanobacterium *Prochloron didemni* [14] and the intracellular “*Candidatus* Endolissoclinum faulkneri” [15]. The former is capable of horizontal transfer, as evidenced by its almost clonal genome amongst very divergent hosts and lack of genome reduction [14], while the latter is vertically transmitted, as evidenced by its divergence with its hosts across cryptic speciations, and profound genome reduction [15, 16]. However, the mode of transmission also exists on a continuum from strict vertical to strict horizontal, with mixtures of vertical and horizontal transmission in between. For instance, the Tsetse fly symbiont *Sodalis*

glossinidius shows some signs of genome-reduction such as rampant pseudogenes, but remains culturable in the lab, meaning that horizontal transfer cannot be excluded [17]. Likewise, symbionts long thought to be exclusively vertically transmitted, such as the bryozoan symbiont "*Ca.* Endobugula sertula", which is packaged with the hosts' larvae, show no signs of genome reduction [18], meaning that there is no compelling reason why it should not be able to transfer horizontally between hosts. Indeed, "*Ca*. E. sertula" has been found in genetically divergent but proximal bryozoan individuals, suggesting horizontal transfer [19].

The *Lagria* and *Ecnolagria* beetles belong to subfamily Lagriinae within the family Tenebrionidae (order Coleoptera). The *Lagria villosa* beetle, a known soybean pest [20], is a source of lagriamide, an antifungal polyketide [20]. The compound is produced via a *trans*-AT polyketide synthase (PKS)-non-ribosomal peptide synthetase (NRPS) hybrid biosynthetic gene cluster (BGC), termed *lga,* which due to a nucleotide signature (kmer frequency) distinct from the chromosome is predicted to have been horizontally acquired [20]. The *lga* BGC is encoded by a *Burkholderia* symbiont (*Burkholderia* sp. LvStB) that is present in glandular structures associated with the ovipositor of female beetles and secreted on its eggs as they are laid [20]. The symbiont has been shown to have a defensive role against fungi in the egg [20] and larval stages [21]. Previously, we showed that the genome of *Burkholderia* sp. LvStB is reduced and has lost a number of essential genes including some genes involved in the DNA repair pathways and primary metabolism [22]. The genome has a low coding density, and a high number of pseudogenes and transposases, indicative of a reduced genome [22]. These characteristics are consistent with host restriction and vertical transmission of LvStB. However, there is evidence that *Burkholderia* symbionts from *L. villosa* can be transferred to plant tissues and persist for several days, and that bacteria can be acquired by the beetle from the plant and soil environment [23].

As some Lagriinae beetles harbour symbionts in special structures that likely evolved between 55 and 82 million years ago based on fossil evidence, positioned to deposit symbionts on the eggs [24, 25], we hypothesised that lagriamide-producing *Burkholderia* symbionts might have co-evolved with their hosts in a manner similar to other vertically-transmitted insect symbionts. However, the possibility for transmission of the symbionts to and from plants, and the accessibility of the symbionts' habitat on the surface of eggs and within adult females suggested that horizontal symbiont acquisition may be possible. As the beetles harbour complex microbiomes with multiple related *Burkholderia* strains as well as other bacteria [21, 23], both genome-reduced and not, an alternative hypothesis is that the lagriamide BGC has been repeatedly horizontally transferred among environmental strains and symbionts. Moreover, partnerships in defensive symbionts are usually more dynamic as compared to intracellular nutritional symbionts [26]. It's also possible that the lagriamide-producing strain is restricted to *L. villosa*, and that different Lagriinae species have symbionts with different BGCs, as this would allow the association to react much more flexibly to changes in antagonist communities. To clarify this evolutionary picture, we analysed the metagenome of 12 beetle populations, spanning eight species belonging to genera *Lagria* and *Ecnolagria* across five different countries (four continents) (**Table 1**). We recovered the metagenome-assembled genomes (MAGs) of several different *Burkholderia* bacteria, and confirmed the presence of the lagriamide BGC in each. We also report a complete genome of the genome-reduced, lagriamide-producing *Burkholderia* sp. LvStB symbiont, obtained through long-read Nanopore sequencing. We compared the phylogeny of the recovered *Burkholderia* MAGs, the lagriamide BGCs, and the host beetles to determine whether co-cladogenesis occurred in this system, and to further explore the evolutionary relationships in the symbiosis. The results indicate that the lagriamide BGC was likely only acquired once in the common ancestor of beetle-associated *Burkholderia* symbionts, and subsequently lost in the majority of the descendent free-living strains. Remarkably, as all the lagriamide-bearing symbionts are genome-reduced but they do not form

a monophyletic clade and do not correspond to host phylogeny, they may represent multiple acquisition events, followed by independent genome-reduction processes. The common factor of lagriamide production might be one of the reasons for selection by and dependency on hosts. This would suggest that a single group of natural products caused many independent symbioses to be established over evolutionary time.

Results and Discussion

Beetle phylogeny

We sequenced and assembled the metagenomes of the collected *Lagria* and *Ecnolagria* beetle populations (**Table 1**), and beetle mitogenomes were extracted and annotated to infer host beetle phylogeny (**Fig. 1**). In line with previous studies, mitogenomes belonging to subfamilies Lagriinae, Blaptinae, Pimeliinae, Stenochiinae and Alleculinae were found to be monophyletic whereas Diaperinae and Tenebrioninae were found to be para- or polyphyletic [27–30]. Maximum likelihood analysis using RAxML [31] (**Fig. SI 1**) and Bayesian analysis using MrBayes [32] (**Fig. 1**) gave the same results.

All collected *Lagria* beetle mitogenomes clustered into four distinct subclades: All *L. hirta* beetle mitogenomes were clustered in a single clade, alongside a closely related clade of several *Lagria* species (*L. rufipennis*, *L. okinawana* and *L. nigricollis*) from Japan. The *L. atripes* and *L. grenieri* beetles formed another clade more distantly related to the *L. hirta* and Japanese *Lagria* species. Finally, the *L. villosa* and *Ecnolagria* sp. beetles formed a fourth clade along with *Chrysolagria* sp. (JX412760), distinct from the other *Lagria* beetles. Furthermore, while publicly available sequences of *L. hirta* (OX375806) clustered with collected *L. hirta* samples from

Rhineland Palatinate, Germany (LhHG), *L. rufipennis* (MW802588) clustered outside the *L. rufipennis* (Lruf), and *L. nigricollis* (Lnig) clade. This slightly distant clustering of the two *L. rufipennis* mitogenomes could be due to the different geographical origins (China and Japan, respectively). Unfortunately, however, the scarcity of taxonomic work and lack of molecular barcodes on the Asian *Lagria* species in the BOLD database [33] preclude us from unambiguously resolving this issue.

Recovery of Lagriamide BGCs

A complete, or mostly complete, *lga* BGC was found in eleven of the twelve samples, with the exception of *L. rufipennis* where only small fragments of the *lga* BGC could be recovered. The BGC recovered from *L. nigricollis* was found over two contigs and could not be manually joined following inspection of the assembly graph. The missing data for this region spans from approximately halfway through the *lgaB* gene to approximately halfway through the *lgaC* gene (**Fig. 2A**). Inspection of the assembly graphs with focus on the *lga* BGC regions revealed several possible variants of the BGC present in the symbionts associated with *L. atripes*, *L. hirta* SB, *L. okinawana*, and *L. nigricollis* (See Supplementary Material). Close analysis of the variants revealed that neither the gene nor domain organisation and predicted functionality thereof was affected within the variants (i.e intra-variant organisation). We used read coverage as a proxy for the most dominant BGC variant and the variant with the greatest coverage support was chosen as the representative BGC for subsequent analyses. Full analysis regarding BGC variants and selection of representative BGCs can be found in Supplementary Material.

Analysis of representative BGCs revealed several differences in gene organisation of the *lga* BGC across the different Lagriinae beetle species (**Fig. 2A**). The first difference we observed was among the *lga* BGCs recovered from the *L. hirta* samples, wherein most *lga* BGCs exhibited a split in *lgaC* (**Fig. 2A**). We initially had only short-read sequence data for the *L. hirta* HG population, from which we recovered a *lga* BGC that exhibited a split in *lgaC*, as observed in other *L. hirta*-derived *lga* BGCs (**Fig. 2A**). However, we later acquired additional long-read sequence data for the *L. hirta* HG population and, following reassembly, were surprised to find that the recovered *lga* BGC did not exhibit the previously observed break in *lgaC*. PCR amplification of this region in the *lgaC* gene revealed that both complete and split variations of *lgaC* were real and present in the *L. hirta* HG population as sequence variants. The coverage of the *lga* BGC is the same as that of the predicted host genome, which would suggest that it is present as a single copy. We therefore concluded that different BGC variants exist in the *L. hirta* HG-associated *Burkholderia* population and may exist in the other Lagriinae populations. Both read coverage and PCR amplification suggest that the variant with the complete (not split) *lgaC* is more abundant.

The second difference we observed was in the *lga* BGC from the LvStB genome extracted from the 2023 *L. villosa* metagenome (Lv23). One split in *lgaB* and two splits in *lgaC* were seen. However, the assembly of the *L. villosa* 2023 metagenome was based solely on long-read data, which is error-prone [34], and the splits may not be a true reflection of the BGC in this sample. Normally Sanger sequencing would be the solution to validate these questionable regions but unfortunately, there was no remaining DNA after the long read sequencing runs for this particular sample. For this reason, we left the BGC with the splits but were cautious not to over-interpret the apparent breaks in the genes in this BGC.

We then considered the domain organisation within the *lga* BGC genes (**Fig. 2B**). The domain organisation is largely congruent across the *lga* BGCs recovered from the metagenomes. We

did note, however, an additional annotated "DHt" domain in *lgaB*, which is defined as "Dehydratase domain variant more commonly found in *trans*-AT PKS clusters", in the *lga* BGCs from all *L. hirta* samples and the *L. nigricollis* population. Similarly, we detected an additional carrier protein domain (phosphopantetheine acyl carrier protein group) near the N-terminus of the *lgaC* protein in the BGCs from the *L. grenieri* and *L. okinawana* samples. In all cases, close inspection of the primary sequence of these additional dehydratase and carrier domains revealed mutations in the sequences that would likely render the enzyme non-functional (see Supplementary Material for full details).

Finally, as with the originally described *lga* BGC recovered from the *L. villosa* 2019 sample [20], we found mutations in the catalytic or conserved motifs of *lgaG* DH2, *lgaG* KS6, *lgaB* KR3 and *lgaC* KS5 domains, that we believe may render these domains inactive (see Supplementary Material for full details). As a result, the domain architecture of all representative *lga* BGCs from all samples are likely functionally identical.

Together, the conservation of the *lga* BGC in at least seven different species of Lagriinae beetles, across four geographically distant countries, implies that the production of lagriamide is an important factor for the host beetle and that the *lga* BGC is under strong selective pressure. The presence of additional domains in the *lga* BGC in several samples, even though they are likely inactive, is intriguing as it suggests that these domains may have previously been present in all *lga* BGCs but may have decayed over time and were lost. The reason as to why these domains were selected against would be speculative at best and inspection of all lagriamide-like compounds produced in the different beetle populations would need to be assessed to truly infer differences that the domain architecture may have on the resulting chemistry. Conserved production of of other bioactive compounds has been observed, such as pederin, across Staphylinidae beetle species (*Paederus* and *Paederidus* genera) [35], which are host to a

Pseudomonas symbiont that produces pederin [36]. Pederin is a toxic compound, produced via a *trans*-AT PKS-NRPS hybrid BGC by the *Pseudomonas* symbiont that is coated onto the beetle eggs and subsequently maintained by the larvae into adulthood and serves to protect them against predation [37].

The two systems have several parallels: both pederin and lagriamide are produced by a *trans*-AT PKS NRPS hybrid BGC in *Pseudomonas* bacterium [35], where the compound is concentrated in the female oviposition organs, coated onto the eggs and serves to protect juveniles [37]. Further, both pederin and lagriamide are the sole insect-associated compounds in suites of compounds otherwise associated with marine invertebrates. Groups of pederin analogs, such as the onnamides, mycalamides, psymberins, and theopederins have been isolated from a variety of marine sponges [38–42] and ascidians [43], while bistramide, the most structurally similar compound to lagriamide, was isolated from an ascidian [44]. Another analogous system is that of beewolf wasps and *Streptomyces* bacteria which exhibit some characteristics of genome erosion [45] and produce at least two classes of defensive compounds: piericidins, and streptochlorins [46]. In this system, the *Streptomyces* bacteria are extracellular symbionts stored in cuticular invaginations in the female wasp antennae and are transmitted to the brood cells and subsequently transferred to the cocoon to protect the larvae from fungal infection [46]. Piericidins are similarly produced by a *Streptomyces* harboured within marine ascidian *Didemnum molle* [47, 48], and streptochlorins were originally isolated from a marine-sediment borne *Streptomyces* species [49]. The question remains, however, as to what the fundamental link is between these terrestrial and marine systems. What is clear, however, is that the production of these compounds is crucial in the presence of predators, which provides additional evidence that the production of lagriamide likely provides equally crucial defence for the Lagriinae beetle eggs and larvae against fungal pathogens [20].

Complete genome of the *lga*-carrying LvStB symbiont

Long read sequencing of the *L. villosa* 2023 (Lv23) metagenome allowed us to assemble a complete genome of a *lga*-carrying *Burkholderia* strain (referred to as LvStB_2023 from hereon). LvStB_2023 was found to have a 2.5 Mbp long genome with a GC percentage of 58.63%. It has 2 circular chromosomes - chromosome 1 is 1.89 Mbp, chromosome 2 is 0.55 Mpb in size, and there is a plasmid 59.77 kbp long. The genome is estimated to be 97.1% complete (98.8% with “specific” mode) and 0.02% contaminated as per CheckM2 [50] and thus is a high quality MAG as per the MIMAG standards [51]. Assembly graph analysis of LvStB_2023 verified that we have the complete sequence of two circular chromosomes and a plasmid. However, the CheckM2 estimate did not reflect a fully complete genome, at 98.8%, and we believe that this small discrepancy in predicted completeness may be a result of ongoing genome reduction [52].

LvStB_2023 has a coding density of 78% and 59.1% with and without pseudogenes respectively. A large percentage (43.87%) of the ORFs in LvStB_2023 were identified as pseudogenes (1613 out of 3676), the highest of any *lga*-carrying *Burkholderia* symbiont. However, this estimate may be artificially high as pseudogenes are identified purely based on their length relative to their closest BLASTP match and these counts are derived from an assembly generated from only long-read data which can be prone to errors, particularly homopolymeric runs. However, coding density and frequency of pseudogenes is not a great deal different to LvStB MAGs assembled from short-read data (see Table S1 for complete genome characteristics of recovered MAGs). Having multiple chromosomes is a common phenomenon in *Burkholderia* [53, 54]. Generally in multi-chromosome bacteria, the majority of the genes for essential functions are located on one larger or primary chromosome. The smaller or the secondary chromosome has much fewer essential genes and mostly carries genes for niche specific functions [55]. In the case of LvStB_2023, chromosome 1 appears to be the

primary chromosome as it is much larger in size, and has 77 out of 84 core genes (including multiple copies) (**Fig 3A**). Functional analysis revealed chromosome 1 to have the highest number of genes for all essential COG categories (**Fig 3B**), including categories L (replication, recombination and repair), J (Translation, ribosomal structure and biogenesis), M (Cell wall/membrane/envelope biogenesis) and H (Coenzyme transport and metabolism).

The *lga* BGC is on chromosome 2 (0.55 Mbp long) and can be distinguished by the continuous block of reverse coding sequences in **Fig 3A**. A detailed view of the *lga*-carrying chromosome can be seen in **Fig. SI 2,** where the GC content of the *lga* BGC is significantly higher than the rest of the genome, providing further evidence towards its horizontal acquisition. Chromosome 1, chromosome 2 and the plasmid have 44.75%, 37.59% and 42.34% of their coding capacity taken up by pseudogenes, respectively. The similar abundance of pseudogenes in each of the contigs indicates that the whole genome is undergoing reduction simultaneously. The chromosome with *lga* (chromosome 2) has the smallest percentage of pseudogenes, which may be a reflection of the required conservation of the *lga* BGC in combination with the presence of large genes in *lga*.

Diversity of beetle-associated *Burkholderia* symbionts

Recovery and analysis of metagenome assembled genomes

Following assembly, the 12 beetle metagenomes were binned, and resultant bins manually refined. A total of 77 MAGs were recovered from all samples, of which 24 MAGs were of high quality, 30 of medium quality, and 23 of low quality (**Table S1**) in accordance with published MIMAG standards [51]. Only medium and high quality MAGs were used for downstream analysis, with the exception of one low quality bin carrying the *lga* BGC (LhHG_2). Confirmation

that the *lga* BGC had been assigned to the correct MAG has been covered in detail in the Supplementary Material. Genome erosion, such as that already observed for the *lga*-carrying symbiont *Burkholderia* sp. LvStB [22], can skew the completeness metric. To determine if a lower quality MAG was incomplete or genome-reduced, we also considered several other metrics, including core gene presence, number of pseudogenes [22], and coding density (**Table S1**), and concluded that this particular MAG (LhHG_2) was likely both reduced and incomplete.

For each beetle population a single MAG belonging to genus *Burkholderia* with a single copy of the *lga* BGC was identified, with the exception of the symbiont within *L. nigricollis* where the possibility exists that the *lga* BGC is present in two copies (Please see Supplementary Material for full details). Previous studies into the lagriamide-carrying symbiont strain *B. gladioli* LvStB [20, 22, 56], showed that this strain was significantly more abundant than all other bacteria associated with *L. villosa*, and had a reduced genome. Consistent with this, all newly recovered MAGs that included *lga* BGCs were the most abundant MAGs in each sample and had reduced genomes with an abundance of pseudogenes and transposases, and had lower coding densities relative to other *B. gladioli* genomes (**Table S1**). In standing with previous studies of *Lagria* beetles, where both reduced and non-reduced *B. gladioli* genomes were recovered, additional *B. gladioli* MAGs (Latri_2, LhHG_3, and LhSB_5) were recovered that did not carry the lagriamide BGC and showed no evidence of genome erosion. We also recovered three small *B. gladioli* MAGs (Lgren_7, Lv19_6_18, Lv20_2) and one small *Burkholderia* MAG (Lv19_6_14). We also recovered MAGs classified as *B. lata* (Lv19_4_0) and *B. arboris* (Lv20_1).

Average nucleotide identity (ANI) analysis of *B. gladioli* MAGs carrying the *lga* BGC showed that MAGs from different beetle species and/or different locations were likely different bacterial species due to shared ANI values less than 95% [57]. Studies have suggested that ANI alone is not a sufficient metric for species delineation and that the aligned fraction (AF) must also be

taken into account [57–60]. Following recent cutoffs adopted for species delineation [58], we opted to use AF ≥ 60% along with ANI ≥ 95% as a cutoff for species assignment. Subsequently, we found that the *Burkholderia* MAGs carrying the lagriamide BGC appeared to be split into at least five novel species (**Table S2**).

Phylogenetic analysis of recovered metagenome assembled genomes

Phylogeny of the *lga*-carrying *Burkholderia* symbionts, relative to other *Burkholderia* species, was inferred using hierarchical orthogroups identified using Orthofinder [61] (**Fig 4** and **Fig. SI 3**). *Burkholderia* symbionts with the *lga* BGC were closely related but did not form a monophyletic clade. Additionally, we included the genome of a free-living *Paraburkholderia acidicola* ATCC 31363 from soil which was recently found to carry a related lagriamide-like BGC (*lgb*), that encodes a structurally similar compound [62] (**Fig SI 4**). This bacterium clustered with other *Paraburkholderia* species.

Unexpectedly, the *Burkholderia* symbionts with a *lga* BGC did not form a monophyletic clade, as would be expected in a long-term vertically-transmitted symbiont, and their common ancestor existed before the common ancestor of many of the free-living *B. gladioli* strains. Given that it is highly unlikely that genome-eroded symbionts could revert to free-living strains with large genomes, the symbiont phylogeny is indicative of at least four independent transitions to a symbiotic lifestyle in closely related bacteria with the *lga* BGC. Each of the transitions was subsequently followed by genome erosion. The lack of synteny observed in the genes flanking the *lga* BGC is indicative of either genomic rearrangement or independent acquisition of the *lga*

BGC, and further supports the independent acquisitions of symbionts followed by genome erosion (**Fig SI 5**).

The phylogeny of the *lga* BGC-carrying *Burkholderia* symbionts was found to be largely incongruent with the beetle phylogeny (**Fig 5A**). The only clades that showed possible co-cladogenesis with the beetle host were the clade of *L. hirta* and their associated BGCs with the clade of *L. nigricollis* and *L. rufipennis*, and their associated *lga* BGCs. This incongruence in the symbiont-host phylogeny was indicative to us of possible host jumping of the symbiont or on-going replacement of the symbiont due to sporadic loss from genomic degradation of the symbiont. Vertical transmission of symbionts is often linked with genome erosion including high numbers of pseudogenes, low coding density and loss of genes required for host-independent survival [6, 63]. Symbiont replacement has often been reported in nutritional symbionts as a way for the hosts to replace a genetically degraded symbiont with a more complete and effective one and to acquire new adaptations for expanding into different niches [64]. *Burkholderia* symbionts related to *B. gladioli* in *Lagria* beetles have been reported to evolve from plant associated bacteria [25] capable of transfer from beetles to plants with subsequent survival [23]. It is possible that the horizontal acquisition might occur in the egg and larval stages, where the symbionts are localised on the surface (eggs) or in cuticular invaginations (larvae and pupae) that remain connected to the external surface via a small duct [65]. As the closely related *Burkholderia* strain LvStA can be acquired horizontally from the environment [23], and there is evidence of free-living bacteria carrying lagriamide-like BGCs [62], we propose that there are *lga*-carrying *Burkholderia* strains persevering in the environment (e.g. in plants or soil) [23] that can be horizontally acquired by the beetle host.

As we previously observed that *lga* has distinct nucleotide composition to the Lv19 genome [20], suggestive of a recent horizontal transfer, we sought to determine if it has been independently transferred to the corresponding symbiont in different beetle hosts. Phylogenetic analysis of the representative *lga* BGCs from all samples resulted in two possible topologies using GTRCAT-V and GTRGAMMAI models (**Fig. 5B**). Both topologies included conserved clades. However, the relative positions of the three clades are poorly supported (**Fig. 5B**, highlighted in blue), resulting in the two alternative topologies. A Bayesian tree was also constructed (**Fig. SI 6**) which is congruent with the GTRGAMMAI tree topology. The inconsistent topology likely stems from limited resolution of the phylogeny affecting deep nodes in the trees, and we interpret the totality of trees to be largely congruent with symbiont phylogeny without additional evidence of more horizontal transfer events. It is likely that there was a single acquisition of *lga* in the common ancestor of the symbiont and *B. gladioli* clade, with subsequent loss in the free-living group. Comparison of both BGC phylogenetic trees with the beetle host phylogeny revealed that both possibilities were incongruent with the beetle host phylogeny (**Fig. 5C**), as seen with the symbionts. It appears that lagriamide production was highly selected for in symbiotic settings and hence retained, whereas it was lost in the most of the larger genomes (assumed to be free-living) where it was not selected for. These findings indicate that the *lga* BGC is important in the symbiosis, either for symbiont establishment (e.g. competition with other symbionts) and/or because lagriamide is an effective host-defensive molecule. Furthemore, the fact that highly similar *Burkholderia* species with *lga* were identified across different Lagriinae beetles indicates that symbiont acquisition is highly selective.

We performed an analysis of pentanucleotide (5-mer) frequency of the beetle-associated, *lga*-carrying symbionts and their associated BGCs, along with the genomes of recently identified soil-borne *Paraburkholderia* species that carry the lagriamide B (*lgb)* BGC, which is highly similar to the *lga* BGC [62], to determine if there was any evidence that the *lga* BGC had been

403 originally acquired from an organism like the *Paraburkholderia*. Visualisation of 5-mer
404 frequencies of the BGCs and the genomes revealed three clusters of BGCs: The BGCs from the
405 two soil-borne *Paraburkholderia* strains, the BGCs from the Brazilian *L. villosa*-derived LvStB
406 strains, and then a third cluster of all other *lga* BGCs (**Fig. SI 7**). A similar pattern was observed
407 for the nucleotide composition of the respective genomes wherein LvStB and Lv20_9 form an
408 isolated cluster, the two soil-borne *Paraburkholderia* form a second, distant cluster, and all other
409 *lga*-carrying Burkholderia strains and cultured *Lagria*-associated genomes (LvStA and LhStG)
410 form a third cluster. None of the BGCs share similar 5-mer composition with their respective
411 genomes suggesting that all are likely horizontally acquired from an independent donor
412 organism.

413 We noted in a previous analysis of the COG annotated genes that there appeared to be a
414 particularly high number of genes in the L category (replication, recombination and repair) that
415 were likely pseudogenes (**Fig 3B**). We assessed the percentage change of COG annotated
416 genes in all *lga*-carrying *Burkholderia* and found that this pseudogenization of genes involved in
417 DNA replication, recombination and repair was particularly high in all the Brazilian *L. villosa*-
418 derived LvStB strains, as well as MAGs LhSB_1, LhG_1, Loki_2 and Lgren_6 (**Fig. SI 8**). Two
419 of the three the LvStB strains also exhibited high pseudogenization of the genes associated with
420 cell motility (Category N). While COG annotation of genes does not provide a robust picture, as
421 not all genes are successfully annotated, the increased pseudogenization of genes involved in
422 DNA replication and repair may explain the drift of the LvStB strains observed in both the
423 phylogenetic analysis and the related 5-mer analysis. In particular, LvStB MAGs possessed
424 highly truncated and psuedogenized *polA* genes, coding for DNA polymerase I used in many
425 DNA-repair pathways and chromosome replication [66], whereas other *lga*-containing MAGs,
426 except LhHG_2, had intact *polA* genes (**Fig. SI 9**). The loss of *polA* in the *L. villosa* symbionts
427 explains their accelerated sequence drift in the genome as a whole and also in the *lga* BGC

compared to the *lga*-possessing symbionts (**Fig. SI 7**). The relative shortness of the LhHG_2 branch in the symbiont tree suggests that its *polA* gene might have degraded relatively recently.

The presence of the *lga* BGC in closely related *Burkholderia* symbionts and not in free living *B. gladioli* indicates it is heavily selected for in the beetle symbiosis ecosystem and only specific bacteria with the BGC are acquired by the beetles. It is possible that lagriamide might be protective against natural enemies other than fungi. How and why it is highly selected for despite the different geographical locations of these hosts is intriguing and suggests a number of interesting questions, considering that the beetles studied here have access to many other symbionts that can produce secondary metabolites, including antifungals [22], such as the antifungal caryoynencin, which was isolated from the culturable beetle symbiont *B. gladioli* LvStA.

Previous studies have highlighted how symbionts can be conserved across host-speciation events and millions of years, leading to genome reduction in the symbiont [6, 16]. A disadvantage of such an exclusive relationship is that the symbiont inevitably suffers from increasingly severe genome erosion, and the end point of that process may be complete dysfunction [67]. Our data concerning the Lagriinae beetles presented here suggests another possibility, that if there is a reservoir of bacteria in the environment that harbour a pathway for a specific secondary metabolite, they could serve as replacements for degraded symbionts. The apparently repeated acquisitions and then genome reduction of symbionts suggests a unique symbiosis specific to a molecule and not to a symbiont. If that hypothesis is correct, it would suggest that lagriamide plays a role in host-symbiont communication or symbiont acquisition, in addition to being a chemical defence. Perhaps the most intriguing aspect of this system is the seemingly repeated independent genome reduction events in different symbiont clades, which offers a chance to study which aspects of reductive evolution are deterministic and which are stochastic. For instance, the almost exclusive loss of *polA* in LvStB but no other *lga*-containing

453 symbionts except for LhHG_2 suggests that it is a driver of increased sequence drift in that
454 clade. In addition, the selectivity of genome reduction in this group of hosts amongst a complex
455 microbiome suggests a role of the host in selecting for genome reduction, which warrants
456 further study.

457 Methods

458 For full details see Supplementary Material.

459 Insect collection

460 Specimens were collected between 2009 and 2019 in Spain, Germany, Brazil, Japan and
461 Australia in the locations listed in **Table S3**. Female adults were dissected either directly after
462 chilling for ca. 15 min at –20°C or preserved in 70% ethanol or acetone until dissected. The
463 accessory glands were removed and preserved in 70% ethanol at –80°C until further
464 processing. For species in which we suspected the presence of symbiont-harbouring
465 compartments within the ovipositor in addition to the glands, the ovipositor was also dissected
466 and preserved along with the accessory glands.

467 Phylogeny of beetles

468 Mitogenomes were recovered from the Eukaryote kingdom bins from each respective sample.
469 Mitogenomes used by Wei *et al.* [27] in their phylogenetic tree were selected for references and
470 outgroups. Mitochondrial genomes (mitogenomes) from all metagenomic datasets and
471 reference mitogenomes were annotated using the MITOS2 webserver [68] against the RefSe89
472 Metazoan database, using genetic code 5 (invertebrate mitochondrial). Amino acid sequences
473 of the 13 protein coding genes (PCGs) from each mitogenome were collected and aligned using

muscle v5.1 [69]. Nucleic acid sequences of the corresponding PCGs were aligned using pal2nal v14 [70] using the -codontable 5 flag. Nucleic acid alignments were concatenated and a partition file was generated using the pxcat command from the phyx package [71]. Phylogenetic analysis was performed by partitioning each codon position for each gene. Maximum likelihood tree was made using RAxML 8.2.12 (raxmlHPC-PTHREADS-SSE3) [31], with the parameters -f a -m GTRGAMMAI -# 1000 -p 1989 -x 1989. Alignment file in FASTA format was converted to nexus format using Geneious Prime 2023.2.1 (www.geneious.com). Bayesian analysis was performed using MrBayes v3.2 [32] with parameters lset applyto=(all) nst=6 rates=invgamma; and unlink statefreq=(all) revmat=(all) shape=(all) pinvar=(all); using 600,000 generations and sample frequency of 1,000. Phylogenetic trees were visualised and modified in the Interactive Tree of Life server [72].

Lagriamide BGC phylogeny

Lagriamide BGC genes from *lgaA* to *lgaI* were extracted from the representative BGCs. Protein sequences were aligned with muscle v5.1 [69] followed by alignment of DNA sequences using pal2nal v14 [70] using -codontable 11. Subsequent steps were similar to those performed while constructing beetle phylogeny. For RAxML phylogeny using GTRCAT -V, partitioning was only performed per gene as it resulted in higher bootstrap values than partitioning for each codon position in each gene.

Burkholderia symbiont phylogeny

Prokka [73] was used to annotate the ORFs of the genomes. Orthofinder v2.5.4 [61] was run on amino acid sequences of the genomes. A custom script was used to extract the genes with hierarchical orthogroups (HOGs) that are present in more than 90% of the genomes and are only present once per genome (149 HOGs were selected). Muscle v5.1 [69] was used to align

the amino acid sequences of the selected HOGs, followed by pal2nal v14 [70] to align the corresponding nucleic acid sequences using -codontable 11. Subsequent steps were similar to those performed while constructing beetle phylogeny.

For symbiont phylogeny with *P. acidicola* ATCC 31363, 152 HOGs (identified using Orthofinder) present only once and in more than 90% of the genomes were used. Subsequent steps were similar to the symbiont phylogeny generated above were followed and the final tree was generated using RAxML 8.2.12 (raxmlHPC-PTHREADS-SSE3) [31], with the parameters -f a -m GTRGAMMAI -# 1000 -p 1989 -x 1989.

Data availability

The data associated with this study was deposited under BioProject accession no. PRJNA1054523. Metagenomic reads have been deposited in the Sequence Read Archive with accessions SRR27332963–SRR27332975. Representative *lga* BGC sequences have been submitted to Genbank with accession numbers PP034267–PP034279. All lagriamide BGC-carrying MAGs were deposited with the following accession numbers: Ecno_1, JAYFRU000000000; Latri_1, JAYFRV000000000; Lgren_6, JAYFRW000000000; LhG_1, JAYFRX000000000; LhHG_2, JAYFRY000000000; LhSB_1, JAYFRZ000000000; Lnig_2, JAYFSA000000000; Loki_2, JAYFSB000000000; Lruf_1, JAYFSC000000000; Lv20_9, JAYFSD000000000.

Acknowledgments

We would like to thank Jean Keller for providing python scripts that helped in creating the symbiont phylogeny. We are grateful to Rebekka Janke, Takema Fukatsu, Kiyoshi Ando, Kimio

Masumoto, Juan José López and Rolf Beutel for support with insect collection and to Domenica Schnabelrauch for carrying out PCRs, cloning and Sanger sequencing.

Author Contributions

SU: Conceptualization, Data curation, Formal Analysis, Investigation, Methodology, Visualization, Writing – original draft, Writing – review & editing. SCW: Conceptualization, Data curation, Formal Analysis, Investigation, Methodology, Visualization, Writing – original draft, Writing – review & editing. AN: Data curation, Formal Analysis, Investigation, Methodology HV: Data curation, Formal Analysis, Investigation, Methodology. LVF: Conceptualization, Investigation, Writing – review & editing, Supervision, Funding acquisition.
MK: Formal Analysis, Methodology, Conceptualization, Investigation, Writing – review & editing, Supervision, Funding acquisition. JCK: Conceptualization, Investigation, Writing – review & editing, Supervision, Funding acquisition

Funding

This work was supported by the U.S. National Science Foundation [DBI-1845890 to J.C.K., S.U., and S.C.W.].

Competing Interests

The Kwan lab is planning to offer their metagenomic binning pipeline Autometa on the paid bioinformatics and computational platform BatchX in addition to distributing it through open source channels.

References

1. Dyall SD, Brown MT, Johnson PJ. Ancient Invasions: From Endosymbionts to Organelles. *Science* 2004; **304**: 253–257.
2. Stephens TG, Gabr A, Calatrava V, Grossman AR, Bhattacharya D. Why is Primary Endosymbiosis so Rare? *New Phytol* 2021; **231**: 1693–1699.
3. Von Dohlen CD, Moran NA. Molecular Data Support a Rapid Radiation of Aphids in the Cretaceous and Multiple Origins of Host Alternation. *Biol J Linn Soc Lond* 2000; **71**: 689–717.
4. Nováková E, Hypša V, Klein J, Foottit RG, von Dohlen CD, Moran NA. Reconstructing the Phylogeny of Aphids (Hemiptera: Aphididae) Using DNA of the Obligate Symbiont *Buchnera aphidicola*. *Mol Phylogenet Evol* 2013; **68**: 42–54.
5. Chong RA, Park H, Moran NA. Genome Evolution of the Obligate Endosymbiont *Buchnera aphidicola*. *Mol Biol Evol* 2019; **36**: 1481–1489.
6. McCutcheon JP, Moran NA. Extreme Genome Reduction in Symbiotic Bacteria. *Nat Rev Microbiol* 2011; **10**: 13–26.
7. Mira A, Ochman H, Moran NA. Deletional Bias and the Evolution of Bacterial Genomes. *Trends Genet* 2001; **17**: 589–596.
8. Hershberg R, Petrov DA. Evidence that Mutation is Universally Biased Towards AT in Bacteria. *PLoS Genet* 2010; **6**: e1001115.
9. Dietel A-K, Merker H, Kaltenpoth M, Kost C. Selective Advantages Favour High Genomic AT-Contents in Intracellular Elements. *PLoS Genet* 2019; **15**: e1007778.
10. Dufresne A, Garczarek L, Partensky F. Accelerated Evolution Associated with Genome Reduction in a Free-Living Prokaryote. *Genome Biol* 2005; **6**: R14.
11. Morris JJ, Lenski RE, Zinser ER. The Black Queen Hypothesis: Evolution of Dependencies

Through Adaptive Gene Loss. *MBio* 2012; **3**: e00036–12.

12. Pinto-Carbó M, Gademann K, Eberl L, Carlier A. Leaf Nodule Symbiosis: Function and Transmission of Obligate Bacterial Endophytes. *Curr Opin Plant Biol* 2018; **44**: 23–31.
13. Pinto-Carbó M, Sieber S, Dessein S, Wicker T, Verstraete B, Gademann K, et al. Evidence of Horizontal Gene Transfer Between Obligate Leaf Nodule Symbionts. *ISME J* 2016; **10**: 2092–2105.
14. Donia MS, Fricke WF, Partensky F, Cox J, Elshahawi SI, White JR, et al. Complex Microbiome Underlying Secondary and Primary Metabolism in the Tunicate-*Prochloron* Symbiosis. *Proc Natl Acad Sci U S A* 2011; **108**: E1423–32.
15. Kwan JC, Donia MS, Han AW, Hirose E, Haygood MG, Schmidt EW. Genome Streamlining and Chemical Defense in a Coral Reef Symbiosis. *Proc Natl Acad Sci U S A* 2012; **109**: 20655–20660.
16. Kwan JC, Schmidt EW. Bacterial Endosymbiosis in a Chordate Host: Long-Term Co-Evolution and Conservation of Secondary Metabolism. *PLoS One* 2013; **8**: e80822.
17. Toh H, Weiss BL, Perkin SAH, Yamashita A, Oshima K, Hattori M, et al. Massive Genome Erosion and Functional Adaptations Provide Insights Into the Symbiotic Lifestyle of *Sodalis glossinidius* in the Tsetse Host. *Genome Res* 2006; **16**: 149–156.
18. Miller IJ, Vanee N, Fong SS, Lim-Fong GE, Kwan JC. Lack of Overt Genome Reduction in the Bryostatin-Producing Bryozoan Symbiont, '*Candidatus* Endobugula sertula'. *Appl Environ Microbiol* 2016; **82**: 6573–6583.
19. Linneman J, Paulus D, Lim-Fong G, Lopanik NB. Latitudinal Variation of a Defensive Symbiosis in the *Bugula neritina* (Bryozoa) Sibling Species Complex. *PLoS One* 2014; **9**: e108783.
20. Flórez LV, Scherlach K, Miller IJ, Rodrigues A, Kwan JC, Hertweck C, et al. An Antifungal Polyketide Associated With Horizontally Acquired Genes Supports Symbiont-Mediated Defense in *Lagria villosa* Beetles. *Nat Commun* 2018; **9**: 2478.

587 21. Janke RS, Kaftan F, Niehs SP, Scherlach K, Rodrigues A, Svatoš A, et al. Bacterial
588 Ectosymbionts in Cuticular Organs Chemically Protect a Beetle During Molting Stages.
589 *ISME J* 2022; **16**: 2691–2701.
590 22. Waterworth SC, Flórez LV, Rees ER, Hertweck C, Kaltenpoth M, Kwan JC. Horizontal
591 Gene Transfer to a Defensive Symbiont With a Reduced Genome in a Multipartite Beetle
592 Microbiome. *MBio* 2020; **11**: e02430–19.
593 23. Wierz JC, Gaube P, Klebsch D, Kaltenpoth M, Flórez LV. Transmission of Bacterial
594 Symbionts With and Without Genome Erosion Between a Beetle Host and the Plant
595 Environment. *Front Microbiol* 2021; **12**: 715601.
596 24. Jürgen Stammer H. Die Symbiose Der Lagriiden (Coleoptera). *Z Morph u Okol Tiere* 1929;
597 **15**: 1–34.
598 25. Flórez LV, Scherlach K, Gaube P, Ross C, Sitte E, Hermes C, et al. Antibiotic-Producing
599 Symbionts Dynamically Transition Between Plant Pathogenicity and Insect-Defensive
600 Mutualism. *Nat Commun* 2017; **8**: 15172.
601 26. Flórez LV, Biedermann PHW, Engl T, Kaltenpoth M. Defensive Symbioses of Animals With
602 Prokaryotic and Eukaryotic Microorganisms. *Nat Prod Rep* 2015; **32**: 904–936.
603 27. Wei Z, Shi A. The Complete Mitochondrial Genomes of Four Lagriine Species (Coleoptera,
604 Tenebrionidae) and Phylogenetic Relationships Within Tenebrionidae. *PeerJ* 2023; **11**:
605 e15483.
606 28. Wu C, Zhou Y, Tian T, Li T-J, Chen B. First Report of Complete Mitochondrial Genome in
607 the Subfamily Alleculinae and Mitochondrial Genome-Based Phylogenetics in
608 Tenebrionidae (Coleoptera: Tenebrionoidea). *Insect Sci* 2022; **29**: 1226–1238.
609 29. Ragionieri L, Zúñiga-Reinoso Á, Bläser M, Predel R. Phylogenomics of Darkling Beetles
610 (Coleoptera: Tenebrionidae) From the Atacama Desert. *PeerJ* 2023; **11**: e14848.
611 30. Kergoat GJ, Soldati L, Clamens A-L, Jourdan H, Jabbour-Zahab R, Genson G, et al. Higher
612 Level Molecular Phylogeny of Darkling Beetles (Coleoptera: Tenebrionidae). *Syst Entomol*

2014; **39**: 486–499.

31. Stamatakis A. RAxML Version 8: A Tool for Phylogenetic Analysis and Post-analysis of Large Phylogenies. *Bioinformatics* 2014; **30**: 1312–1313.
32. Ronquist F, Teslenko M, van der Mark P, Ayres DL, Darling A, Höhna S, et al. MrBayes 3.2: Efficient Bayesian Phylogenetic Inference and Model Choice Across a Large Model Space. *Syst Biol* 2012; **61**: 539–542.
33. Ratnasingham S, Hebert PDN. BOLD : The Barcode of Life Data System (www.barcodinglife.org). *Mol Ecol Notes* 2007; **7**: 355–364.
34. Hu Y, Fang L, Nicholson C, Wang K. Implications of Error-Prone Long-Read Whole-Genome Shotgun Sequencing on Characterizing Reference Microbiomes. *iScience* 2020; **23**: 101223.
35. Piel J, Höfer I, Hui D. Evidence for a Symbiosis Island Involved in Horizontal Acquisition of Pederin Biosynthetic Capabilities by the Bacterial Symbiont of *Paederus fuscipes* Beetles. *J Bacteriol* 2004; **186**: 1280–1286.
36. Kador M, Horn MA, Dettner K. Novel Oligonucleotide Probes for *in Situ* Detection of Pederin-Producing Endosymbionts of *Paederus riparius* Rove Beetles (Coleoptera: Staphylinidae). *FEMS Microbiol Lett* 2011; **319**: 73–81.
37. Kellner RLL, Dettner K. Differential Efficacy of Toxic Pederin in Deterring Potential Arthropod Predators of *Paederus* (Coleoptera: Staphylinidae) Offspring. *Oecologia* 1996; **107**: 293–300.
38. Witczak ZJ, Bommareddy A, VanWert AL. Pederin, Psymberin and the Structurally Related Mycalamides: Synthetic Aspects and Biological Activities. In: Kim S-K (ed). *Handbook of Anticancer Drugs from Marine Origin*. 2015. Springer, Cham, pp 683–700.
39. Sakemi S, Ichiba T, Kohmoto S, Saucy G, Higa T. Isolation and Structure Elucidation of Onnamide A, a New Bioactive Metabolite of a Marine Sponge, *Theonella* sp. *J Am Chem Soc* 1988; **110**: 4851–4853.

40. Perry NB, Blunt JW, Munro MHG, Thompson AM. Antiviral and Antitumor Agents From a New Zealand Sponge, Mycale sp. 2. Structures and Solution Conformations of Mycalamides A and B. *J Org Chem* 1990; **55**: 223–227.

41. West LM, Northcote PT, Hood KA, Miller JH, Page MJ. Mycalamide D, a New Cytotoxic Amide From the New Zealand Marine Sponge *Mycale* Species. *J Nat Prod* 2000; **63**: 707–709.

42. Simpson JS, Garson MJ, Blunt JW, Munro MH, Hooper JN. Mycalamides C and D, Cytotoxic Compounds From the Marine Sponge *Stylinos* n. Species. *J Nat Prod* 2000; **63**: 704–706.

43. Dyshlovoy SA, Fedorov SN, Kalinovsky AI, Shubina LK, Bokemeyer C, Stonik VA, et al. Mycalamide A Shows Cytotoxic Properties and Prevents EGF-Induced Neoplastic Transformation Through Inhibition of Nuclear Factors. *Mar Drugs* 2012; **10**: 1212–1224.

44. Gouiffes D, Juge M, Grimaud N, Welin L, Sauviat MP, Barbin Y, et al. Bistramide A, a New Toxin From the Urochordata *Lissoclinum bistratum* Sluiter: Isolation and Preliminary Characterization. *Toxicon* 1988; **26**: 1129–1136.

45. Nechitaylo TY, Sandoval-Calderón M, Engl T, Wielsch N, Dunn DM, Goesmann A, et al. Incipient Genome Erosion and Metabolic Streamlining for Antibiotic Production in a Defensive Symbiont. *Proc Natl Acad Sci U S A* 2021; **118**: e2023047118.

46. Engl T, Kroiss J, Kai M, Nechitaylo TY, Svatoš A, Kaltenpoth M. Evolutionary Stability of Antibiotic Protection in a Defensive Symbiosis. *Proc Natl Acad Sci U S A* 2018; **115**: E2020–E2029.

47. Hayakawa Y, Shirasaki S, Shiba S, Kawasaki T, Matsuo Y, Adachi K, et al. Piericidins C_7 and C_8, New Cytotoxic Antibiotics Produced by a Marine *Streptomyces* sp. *J Antibiot* 2007; **60**: 196–200.

48. Hayakawa Y, Shirasaki S, Kawasaki T, Matsuo Y, Adachi K, Shizuri Y. Structures of New Cytotoxic Antibiotics, Piericidins C_7 and C_8. *J Antibiot* 2007; **60**: 201–203.

665 49. Watabe H. A New Antibiotic SF2583A, 4-chloro-5-(3'-indolyl)oxazole, Produced by
666 *Streptomyces*. *Meiji Seika Kenkyu Nenpo* 1988; **27**: 55–62.
667 50. Chklovski A, Parks DH, Woodcroft BJ, Tyson GW. CheckM2: A Rapid, Scalable and
668 Accurate Tool for Assessing Microbial Genome Quality Using Machine Learning. *Nat*
669 *Methods* 2023; **20**: 1203–1212.
670 51. Bowers RM, Kyrpides NC, Stepanauskas R, Harmon-Smith M, Doud D, Reddy TBK, et al.
671 Minimum Information About a Single Amplified Genome (MISAG) and a Metagenome-
672 Assembled Genome (MIMAG) of Bacteria and Archaea. *Nat Biotechnol* 2017; **35**: 725–731.
673 52. Manzano-Marín A, Kvist S, Oceguera-Figueroa A. Evolution of an Alternative Genetic Code
674 in the Providencia Symbiont of the Hematophagous Leech *Haementeria acuecueyetzin*.
675 *Genome Biol Evol* 2023; **15**: evad164.
676 53. Mahenthiralingam E, Urban TA, Goldberg JB. The Multifarious, Multireplicon *Burkholderia*
677 *cepacia* complex. *Nat Rev Microbiol* 2005; **3**: 144–156.
678 54. Lessie TG, Hendrickson W, Manning BD, Devereux R. Genomic Complexity and Plasticity
679 of *Burkholderia cepacia*. *FEMS Microbiol Lett* 1996; **144**: 117–128.
680 55. Egan ES, Fogel MA, Waldor MK. Divided Genomes: Negotiating the Cell Cycle in
681 Prokaryotes With Multiple Chromosomes. *Mol Microbiol* 2005; **56**: 1129–1138.
682 56. Flórez LV, Kaltenpoth M. Symbiont Dynamics and Strain Diversity in the Defensive
683 Mutualism Between *Lagria* Beetles and *Burkholderia*. *Environ Microbiol* 2017; **19**: 3674–
684 3688.
685 57. Goris J, Konstantinidis KT, Klappenbach JA, Coenye T, Vandamme P, Tiedje JM. DNA-
686 DNA Hybridization Values and Their Relationship to Whole-Genome Sequence Similarities.
687 *Int J Syst Evol Microbiol* 2007; **57**: 81–91.
688 58. Varghese NJ, Mukherjee S, Ivanova N, Konstantinidis KT, Mavrommatis K, Kyrpides NC, et
689 al. Microbial Species Delineation Using Whole Genome Sequences. *Nucleic Acids Res*
690 2015; **43**: 6761–6771.

59. Konstantinidis KT, Tiedje JM. Genomic Insights That Advance the Species Definition for Prokaryotes. *Proc Natl Acad Sci U S A* 2005; **102**: 2567–2572.

60. Parks DH, Chuvochina M, Rinke C, Mussig AJ, Chaumeil P-A, Hugenholtz P. GTDB: An Ongoing Census of Bacterial and Archaeal Diversity Through a Phylogenetically Consistent, Rank Normalized and Complete Genome-Based Taxonomy. *Nucleic Acids Res* 2022; **50**: D785–D794.

61. Emms DM, Kelly S. OrthoFinder: Phylogenetic Orthology Inference for Comparative Genomics. *Genome Biol* 2019; **20**: 238.

62. Fergusson CH, Saulog J, Paulo BS, Wilson DM, Liu DY, Morehouse NJ, et al. Discovery of the Polyketide Lagriamide B by Integrated Genome Mining, Isotopic Labeling, and Untargeted Metabolomics. *ChemRxiv* 2023.

63. Lo W-S, Huang Y-Y, Kuo C-H. Winding Paths to Simplicity: Genome Evolution in Facultative Insect Symbionts. *FEMS Microbiol Rev* 2016; **40**: 855–874.

64. Sudakaran S, Kost C, Kaltenpoth M. Symbiont Acquisition and Replacement as a Source of Ecological Innovation. *Trends Microbiol* 2017; **25**: 375–390.

65. Janke RS, Moog S, Weiss B, Kaltenpoth M, Flórez LV. Morphological Adaptation for Ectosymbiont Maintenance and Transmission During Metamorphosis in *Lagria* Beetles. *Front Physiol* 2022; **13**: 979200.

66. Patel PH, Suzuki M, Adman E, Shinkai A, Loeb LA. Prokaryotic DNA Polymerase I: Evolution, Structure, and 'Base Flipping' Mechanism for Nucleotide Selection. *J Mol Biol* 2001; **308**: 823–837.

67. Bennett GM, Moran NA. Heritable Symbiosis: The Advantages and Perils of an Evolutionary Rabbit Hole. *Proc Natl Acad Sci U S A* 2015; **112**: 10169–10176.

68. Donath A, Jühling F, Al-Arab M, Bernhart SH, Reinhardt F, Stadler PF, et al. Improved Annotation of Protein-Coding Genes Boundaries in Metazoan Mitochondrial Genomes. *Nucleic Acids Res* 2019; **47**: 10543–10552.

717 69. Edgar RC. Muscle5: High-Accuracy Alignment Ensembles Enable Unbiased Assessments
718 of Sequence Homology and Phylogeny. *Nat Commun* 2022; **13**: 6968.
719 70. Suyama M, Torrents D, Bork P. PAL2NAL: Robust Conversion of Protein Sequence
720 Alignments Into the Corresponding Codon Alignments. *Nucleic Acids Res* 2006; **34**: W609–
721 W612.
722 71. Brown JW, Walker JF, Smith SA. Phyx: Phylogenetic Tools for Unix. *Bioinformatics* 2017;
723 **33**: 1886–1888.
724 72. Letunic I, Bork P. Interactive Tree of Life (iTOL) v5: An Online Tool for Phylogenetic Tree
725 Display and Annotation. *Nucleic Acids Res* 2021; **49**: W293–W296.
726 73. Seemann T. Prokka: Rapid Prokaryotic Genome Annotation. *Bioinformatics* 2014; **30**:
727 2068–2069.

728

729 **Table 1**. Metadata for different beetles collected for this study. **Lagria villosa* samples were
730 collected three times at different time points. With one sample reported in a previous study
731 (referred to as Lv19 in this work) [20, 22].

Sample	Location	Abbreviation(s) used in the paper
*Lagria villosa**	São Paulo, Brazil	Lv19, Lv20, Lv23
Lagria rufipennis	Osaka and Ibaraki, Japan	Lruf
Lagria okinawana	Okinawa, Japan	Loki
Lagria nigricollis	Tokushima, Osaka, and Kogashima, Japan	Lnig

Lagria hirta	Hessen, Germany	LhSB
Lagria hirta	Rhineland Palatinate, Germany	LhHG
Lagria hirta	Galicia, Spain	LhG
Lagria grenieri	Huelva, Spain	Lgren
Lagria atripes	Rhineland-Palatinate, Germany	Latri
Ecnolagria sp.	New South Wales, Australia	Ecno

732

733 **Figure 1.** Beetle mitogenome phylogenetic tree using 13 mitochondrial protein coding genes
734 constructed using MrBayes [32]. Branch values represent posterior probabilities. Mitogenomes
735 recovered in this study are highlighted with bold lettering.

736

737 **Figure 2.** Analysis of representative *lga* BGCs extracted from eleven Lagriinae beetle
738 metagenomes. A) Comparison of representative *lga* BGC gene organisation. Individual genes in
739 the *lga* BGCs are represented by arrows oriented in the predicted direction of transcription and
740 coloured according to identity. Pairwise amino acid similarity between BGCs is indicated in the
741 shaded areas between genes. A scale bar is provided for gene size. B) Comparison of predicted
742 enzyme domain organisation in the representative *lga* BGCs, where genes are ordered
743 according to biosynthetic order. Boxes around the domains indicate differences between the
744 BGCs.

745

746 **Fig 3**. A) Circular representation of LvStB_2023 genome. B) Raw count of COG categories
747 present on different contigs of LvStB_2023 (with and without pseudogenes).

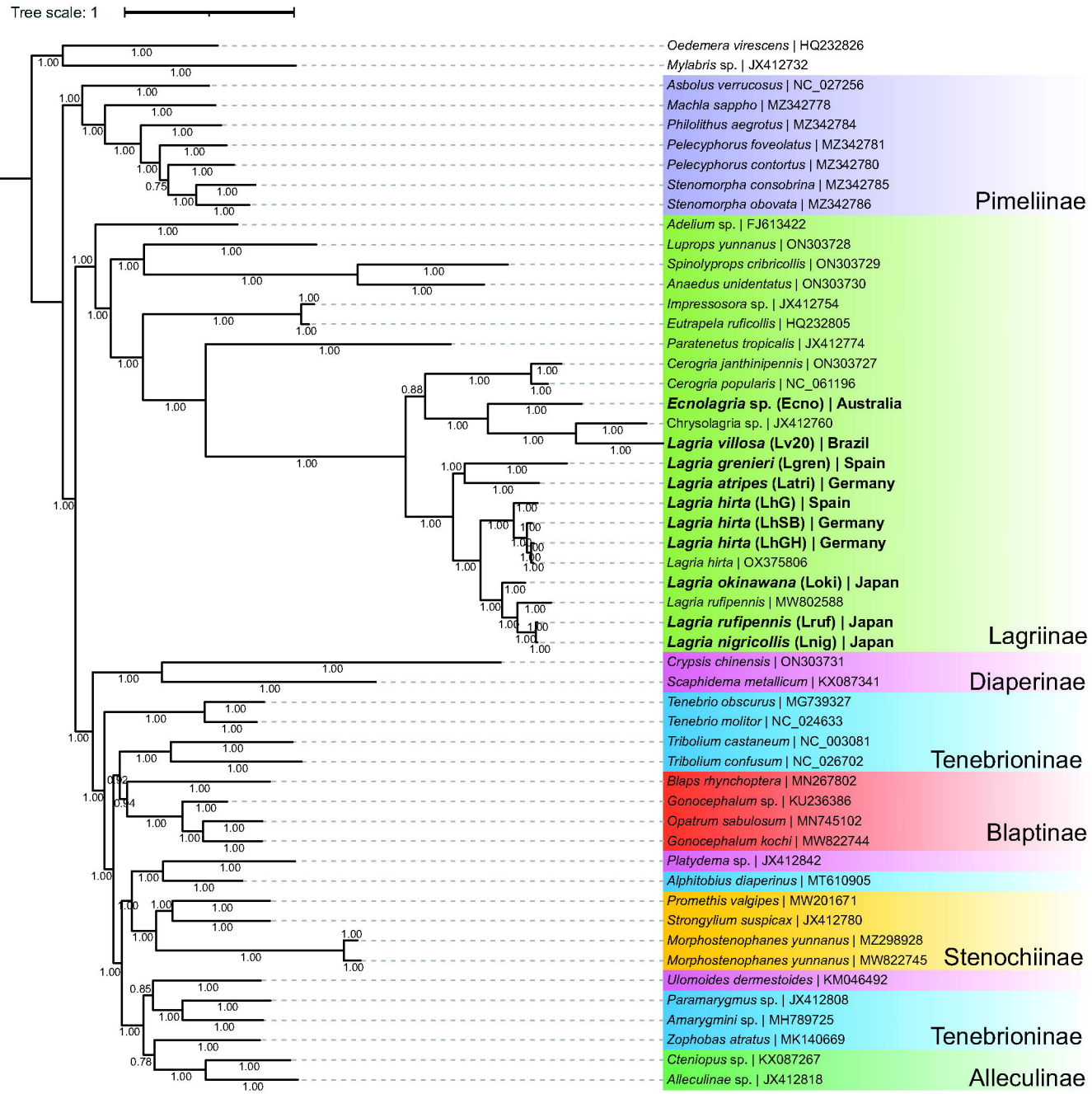
748

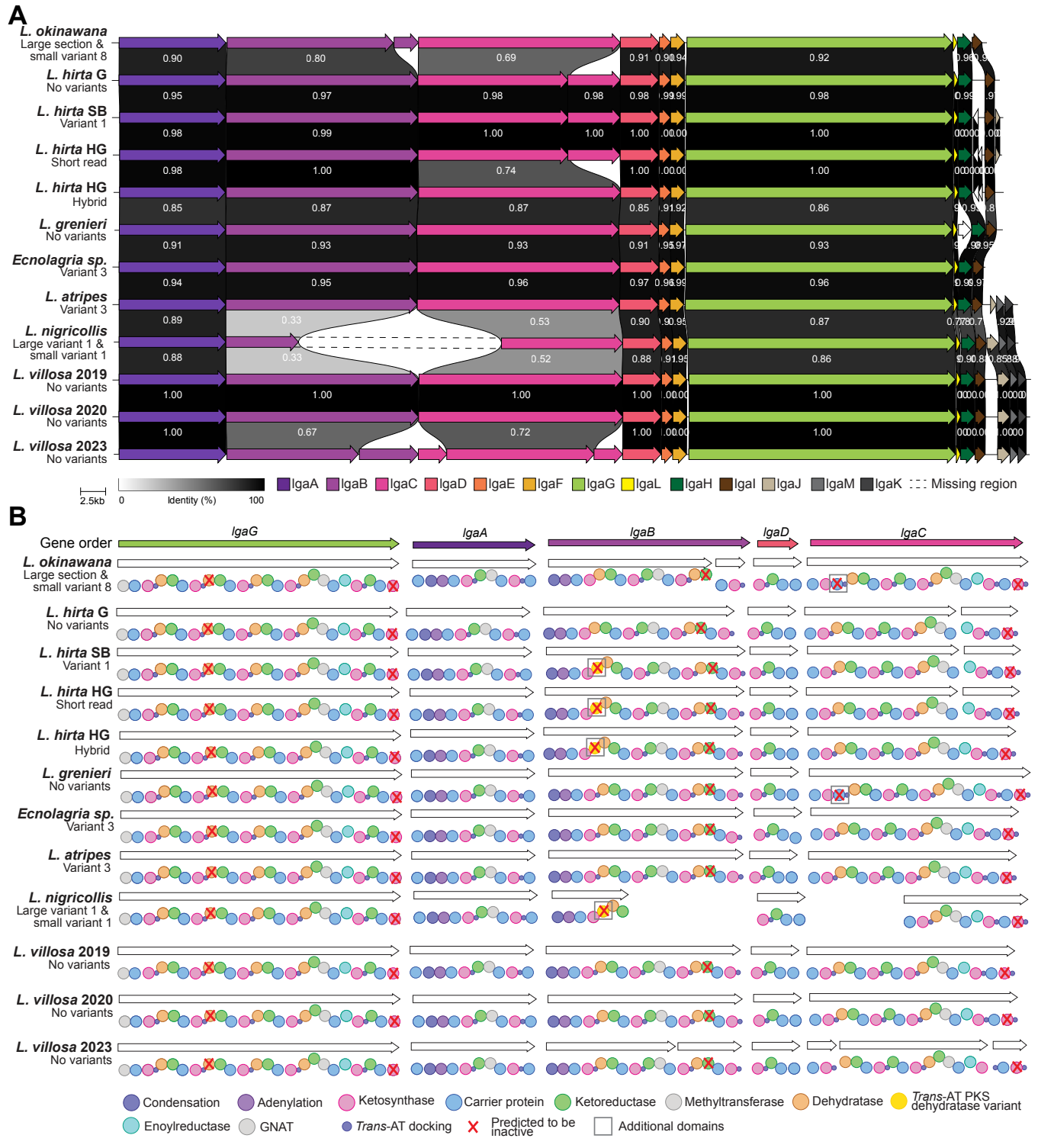
749 **Fig 4**. Bayesian phylogenetic tree of Lagriinae beetle associated *Burkholderia* symbionts.
750 Values on nodes indicate posterior probabilities. Outgroups include - *Paraburkholderia*
751 *acidiphila* (GCF_009789655.1), *Cupriavidus necator* (GCF_000219215.1), *Herbaspirillum*
752 *seropedicae* (GCF_001040945.1).

753

754 **Fig 5**. Congruence between phylogenies of beetle host, *Burkholdria* symbionts and *lga* BGCs in
755 all samples. A) Tanglegram between *lga*-carrying symbionts and beetle host phylogeny. B)
756 Tanglegram between *lga*-carrying symbionts (centre) and the *lga* BGC, as inferred via two
757 models GTRCAT (left) and GTRGAMMAI (right). C) Tanglegram of the BGC phylogenies using
758 the GTRCAT (left) and GTRGAMMAI (right) models relative to the phylogeny of the host beetle
759 based on the respective mitogenome sequences (centre). Clades within the beetle host
760 mitogenome phylogenetic tree are highlighted in shades of grey. In all panels, the three
761 conserved clades are highlighted in purple, green and orange. Dashed brown lines represent
762 nodes that are unique between the respective phylogenies, except when the node is absent
763 from the other tree.

764




A
L. okinawana
Large section & small variant 8
L. hirta G
No variants
L. hirta SB
Variant 1
L. hirta HG
Short read
L. hirta HG
Hybrid
L. grenieri
No variants
Ecnolagria sp.
Variant 3
L. atripes
Variant 3
L. nigricollis
Large variant 1 & small variant 1
L. villosa 2019
No variants
L. villosa 2020
No variants
L. villosa 2023
No variants
2.5kb
0
Identity (%)
100
lgaA lgaB lgaC lgaD lgaE lgaF lgaG lgaL lgaH lgaI lgaJ lgaM lgaK
Missing region
B
Gene order
lgaG
lgaA
lgaB
lgaD
lgaC
L. okinawana
Large section & small variant 8
L. hirta G
No variants
L. hirta SB
Variant 1
L. hirta HG
Short read
L. hirta HG
Hybrid
L. grenieri
No variants
Ecnolagria sp.
Variant 3
L. atripes
Variant 3
L. nigricollis
Large variant 1 & small variant 1
L. villosa 2019
No variants
L. villosa 2020
No variants
L. villosa 2023
No variants
Condensation
Adenylation
Ketosynthase
Carrier protein
Ketoreductase
Methyltransferase
Dehydratase
Trans-AT PKS dehydratase variant
Enoylreductase
GNAT
Trans-AT docking
Predicted to be inactive
Additional domains