Cryptic diversity and signatures of domestication in the Black Soldier Fly (*Hermetia illucens*)

Tomas N. Generalovic[1]*, Christoph Sandrock[2], Benjamin J. Roberts[3], Joana I. Meier[1,4], Martin Hauser[5], Ian A. Warren[1], Miha Pipan[6], Richard Durbin[7] & Chris D. Jiggins[1].

[1]Department of Zoology, University of Cambridge, Cambridge, UK; [2]Department of Livestock Sciences, Research Institute of Organic Agriculture (FiBL), Frick, Switzerland; [3]Georgina Mace Centre for the Living Planet, Faculty of Natural Sciences, Imperial College London, London, UK; [4]Tree of Life Programme, Wellcome Sanger Institute, Wellcome Trust Genome Campus, Hinxton, Cambridge, UK; [5]California Department of Food and Agriculture, Plant Pest Diagnostics Branch, Sacramento, CA, USA; [6]Better Origin, Entomics Biosystems Limited, Cambridge, UK; [7]Department of Genetics, University of Cambridge, Cambridge, UK

*Corresponding Author: Tomas N. Generalovic. Email: tng23@cam.ac.uk

Abstract

The black soldier fly (*Hermetia illucens*) is the prime species in the developing global industry of insects as food and feed, but the genetic basis of its domestication has not been studied. We obtained whole genome sequences for 54 individuals from both wild and captive populations. We identified two genetic lineages at least 3 million years divergent, revealing cryptic diversity within the species complex. Our study indicates that the most common populations used for commercial and academic applications are primarily derived from just one of these divergent lineages, originating from a wild North American progenitor. Despite insect farming practices only reaching mass production scale over the past two decades, we find that captive populations show strong genome-wide signatures of domestication. Selective sweeps are found in multiple independently domesticated populations but are not seen in wild populations and suggest five major domestication loci located on chromosomes two, four and five. These regions contain development, behaviour, reproduction, metabolism and immunity genes. Populations experiencing independent domestication events, in the early stages of domestication, also show evidence of convergent genome evolution with repeated signatures of domestication in these same regions. Some limited evidence of gene flow between divergent lineages was observed, as well as evidence of hybridisation from domesticated populations into the wild. Our study reveals the genetic basis of ongoing domestication and provides a genomic platform for breeding and genetic surveillance in this novel agricultural species.

Keywords: domestication, hybridisation, genetic diversity, insect livestock, cryptic species, selective sweeps, introgression, artificial selection.

Introduction

Domestication of plants and animals has enabled successful human population growth and globalisation (Kareiva et al. 2007). Domestication is a coevolutionary process that emerges from a specialised mutualism where a species fitness is managed through manipulation of the environment, food-source, behaviour, and mate choice (Zeder 2015; Purugganan 2022). Evolutionary changes in physiological, morphological, and behavioural traits over time have ultimately led to specialised varieties of crops and breeds of animal resulting in a host of phenotypic diversity, exemplified by *Brassica oleracea* cultivars and dog breeds (Ostrander et al. 2017; Mabry et al. 2021). This process of domestication provides insights into evolution on short time scales (Darwin 1868; Larson et al. 2014). Whilst domestication has been extensively studied in plants, mammals, birds and fish (Ostrander et al. 2017; Wang et al. 2020; Mabry et al. 2021; Kwon et al. 2022), the silkworm, *Bombyx mori*, is currently the only undisputed domesticated insect (Lecocq et al. 2019). Most livestock exclusively occur as domesticated populations where wild counterparts have become extinct (Sun et al. 2012). Here we focus on the black soldier fly, *Hermetia illucens,* a unique case which persists both in captivity and the wild.

The black soldier fly, *H. illucens* (L., 1758; Diptera: Stratiomyidae), has only become an important and globally mass farmed species within the past two decades (Sheppard et al. 2002; Van Huis 2013; Tomberlin and van Huis 2020). In response to rising food demands driven by an ever-increasing human population, novel and more sustainable agricultural practices are being developed, such as the insect livestock industry (Van Huis et al. 2013; Van Huis 2013). The black soldier fly can convert a diverse range of organic wastes into protein and fat rich biomass making it a key species in the insect-based food and feed industry (Gold et al. 2018; Surendra et al. 2020; Tomberlin and van Huis 2020). The black soldier fly has a cosmopolitan distribution with an origin in the Americas (Marshall et al. 2015), most likely South America (Kaya et al. 2021). Global distribution of *H. illucens* outside of the Americas is predicted to be anthropogenically mediated and by 1960 the species had established most of the present-day distributions (Marshall et al. 2015). Therefore, over 60 years of local adaption to diverse environmental niches has likely occurred. Genomic resources have been established (Zhan et al. 2019; Generalovic et al. 2021) and the species harbours high levels of genetic diversity (Ståhls et al. 2020; Kaya et al. 2021; Guilliet et al. 2022). The large amount of standing genetic diversity suggests considerable potential for artificial selection to develop improved strains (Facchini et al. 2022). However, commercial activity is currently dominated by few closely related breeds all derived from one strain (Kaya et al. 2021). This population was founded in Georgia, USA in the 1990's by Craig Sheppard

with an approximate effective population size of 20,000 individuals (Sheppard et al. 2002; Kaya et al. 2021) and has since been shared across the globe. Captive populations such as these have likely undergone or are currently experiencing a multitude of domestication bottlenecks and genetic adaptions associated with artificial insect farming conditions.

Here we explore *H. illucens* diversity on a whole-genome scale. Using the first genome-wide dataset of *H. illucens* sequenced to high coverage, we explore the genomic landscape of captive and wild populations, from locations across the globe. Using this dataset, we address the following questions: (1) What are the phylogenetic relationships among *Hermetia illucens* lineages? (2) When comparing genomes of captive populations to wild relatives can we reveal common signatures associated with domestication? (3) What is the extent and genetic consequences of introgression in the wild and within insect mass-rearing commercial facilities?

Results

Genome sequencing and variant calling

We sequenced 54 *H. illucens* genomes to an average 22x sequencing coverage alongside an additional outgroup species, *Ptecticus aurifer* (supplementary table 1). In order to capture the extent of global genetic variation within the species we sampled *H. illucens* individuals from genetically distinct (captive and wild) populations, previously identified from microsatellite data (Kaya et al. 2021). The remaining dataset consisted of several representative samples of active industrial and academic captive lines. In addition, we sampled wild sourced populations as part of a taxonomic voucher set. We hereafter term long-established academic/industry-derived samples as ‘captive’ whereas samples directly sampled from in the field were termed ‘wild’. After variant calling and stringent quality filtering, we identified 33,940,069 bi-allelic Single Nucleotide Polymorphisms (SNPs).

Divergent lineages and cryptic species complex

A Maximum-Likelihood (ML) phylogeny of *H. illucens* was reconstructed using *Ptecticus aurifer* as an outgroup. Phylogenetic analysis on the genome-wide SNP data showed two major genetic lineages, the first containing wild sampled individuals from North America and many captives belonging to a single yet diverse sister clade, hereafter named lineage-*a* (Figure 1A). The captive populations found in lineage-*a* are made polyphyletic by the North American wild population. European and North American captive populations form sister clades, or ‘breeds’, in line with a documented split from an original founding population,

named hereafter as the ‘Sheppard’ strain. The ‘EVE’ genome reference line also appears to be a derivative of the Sheppard strain and thus forms a third breed. The second lineage, lineage-*b,* is formed from a diverse range of wild and captive populations. In lineage-*b*, wild and captive clades broadly cluster according to geographic location, indicating these captive populations are commonly established from locally occurring populations and may form new strains. Individuals from Asia (wild) and Australia (captive) also form an ’Australasian’ clade. Phylogenetic methods using coalescence-based methods (ASTRAL) with both Neighbour-Joining (NJ) and ML algorithms also support a split between major lineages (supplementary figure 1A & 1B). However, NJ methods suggest this split is less pronounced with wild western European populations appearing as the most recent common ancestor to the previously identified lineage-*a* (supplementary figure 1A).

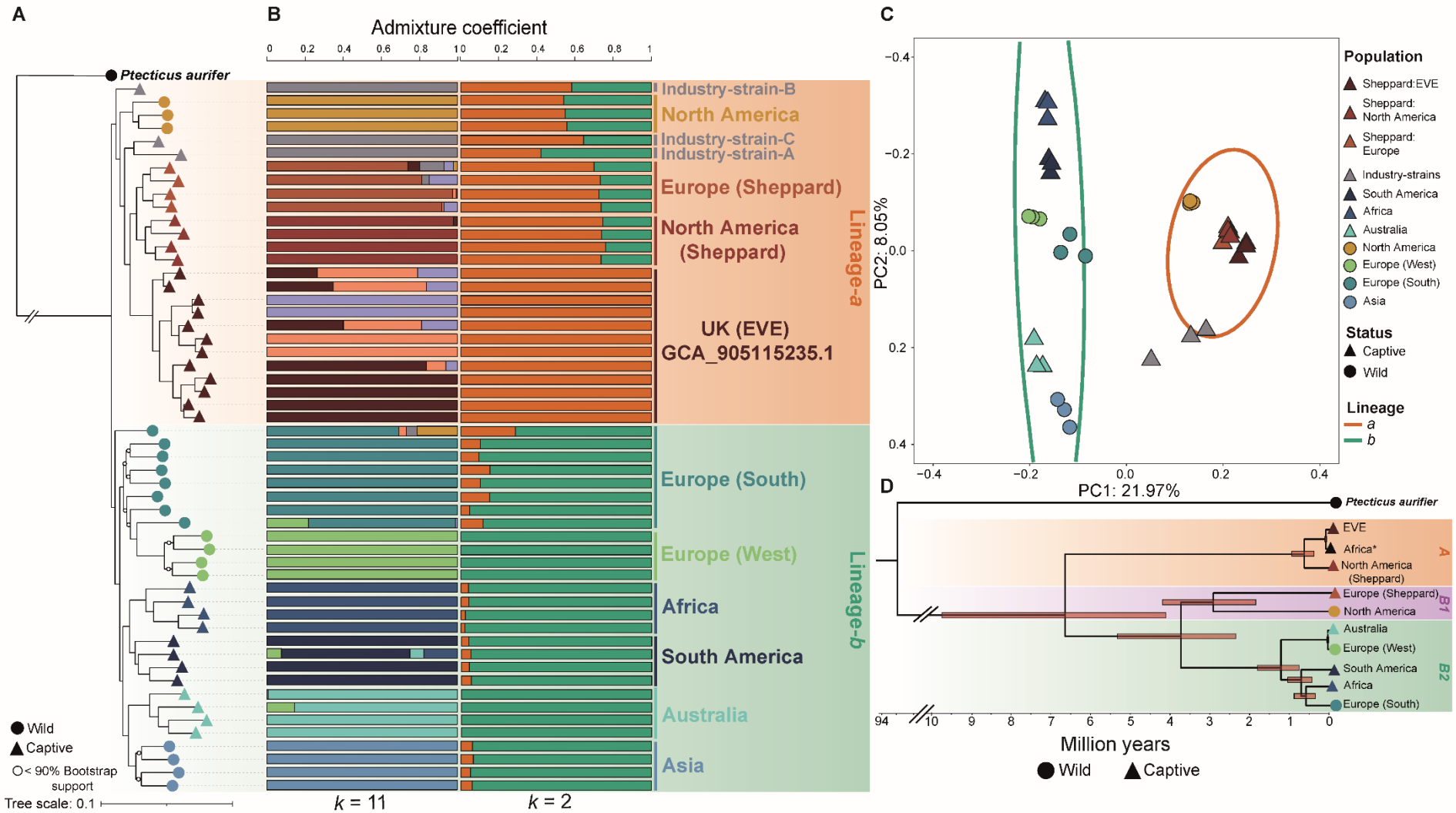

Figure 1. (**A**) Maximum Likelihood phylogenetic reconstruction using whole genomes of *Hermetia illucens* (n=54) and *Ptecticus aurifer* (n=1). (**B**) ADMIXTURE analysis of *H. illucens* population structure at *k*=11 and *k*=2 (lowest cross-validation error support) indicating pronounced global population structure. (**C**) Principal Component Analysis (PCA) using a subsampled dataset of whole genomes (n=3 per population) supporting two discrete genetic lineages. A 95% confidence interval (CI) ellipse is drawn over major lineages based on phylogenetic (**A**) and clustering (**B**; *k*=2) assignment of individuals. (**D**) BEAST time divergence model representation of the *H. illucens* mitochondrial phylogeny reconstructed using a single whole mitochondrial (mt) genome from each major clade identified in the whole-genome phylogeny (**A**). Lineages recovered in the mtDNA tree are labelled including the unresolved mtDNA lineage-*B1*.*Single mt genome obtained from Guilliet et al. (2022).

Evidence for the divergence between the two distinct nuclear lineages was also supported by ADMIXTURE (Figure 1B) and Principal Component Analysis (PCA; Figure 1C). Optimal clustering in ADMIXTURE analysis was identified at k = 2 (Figure 1B; supplementary table 2), and this was consistent when subsampling an equal number of individuals per population (supplementary figure 2). At higher k-values, k = 11 (Figure 1B), wild European populations show distinct clusters indicating a diverse within continent population structure consistent with independent colonisation events and likely subsequent patterns of isolation-by-distance. A similar pattern is observed across a larger geographic range in Asian and Australian samples in agreement with the successive colonisation of Australasia (Kaya et al. 2021; Guilliet et al. 2022). PCA analysis showed that captive *H. illucens* cluster with wild North American samples whilst the remaining wild and captive samples found globally form a second major cluster explained by 22% of the genetic variance along PC1 (Figure 1C). Several individuals belonging to various industry populations (Ind-A, -B & -C) are found in the PCA as intermediates between lineages, suggesting admixed origins for these populations. Measures of genetic differentiation (F_{ST}) and absolute sequence (d_{XY}) divergence were calculated to understand the scale of within and between lineage divergence of *H. illucens*. When grouped by discrete populations identified in clustering analysis we found high genome-wide levels of F_{ST} (0.178) and d_{XY} (0.081) between nuclear lineages. However, we found higher genome-wide values within lineage-*b*, which harbours multiple wild populations from a broad geographic range, as compared to lineage-*a* (supplementary figure 3).

Ancient divergence between *Hermetia illucens* mtDNA lineages

We produced *de novo* mitochondrial assemblies and annotations using whole genome sequence data enriched with mitochondrial DNA (mtDNA) for 54 *H. illucens* and one *P. aurifer* individual (supplementary table 3; supplementary figure 4). The *de novo H. illucens* and *P. aurifer* assemblies are comparable to published reference mitochondrial genomes (supplementary table 3; supplementary figure 5) (Qi et al. 2017; Zhan et al. 2020; Guilliet et al. 2022). Pronounced nuclear lineages (*a* & *b*) were not recovered in the mtDNA tree where three major lineages (*A*, *B1* & *B2*) were identified (Figure 1D & supplementary figure 5). Nuclear lineage-*a* samples were found falling into mitochondria clades *A* and *B1*, which diverged an estimated 6.8 million years ago (my; 95% CI: 4.1 - 9.7 my). Lineage-*B1* of the mtDNA tree contains nuclear lineage-*b* wild North American and captive European-Sheppard populations belonging to nuclear lineage-*a* (Figure 1D), which diverged from mtDNA lineage-*B2* approximately 3.8 my ago (95% CI: 2.4 - 5.3 my). Mitochondrial clade *B2* shows a maximum internal divergence of 1.2 my isolating a captive representative of the

Australasian clade and wild southern Europe from the remaining populations. We also generated Neighbour-Joining (NJ) trees of Cytochrome Oxidase I (COI) markers, such that additional previously published samples could be incorporated, including some from wild Central American populations (supplementary figure 6). This confirms the observation of divergence between two lineages and further supports that Sheppard-derived breeds of lineage-*a* are most closely related to wild samples from North America.

Genetic signals of domestication

Genomic signals of domestication are typically characterised by genome-wide reductions in genetic diversity and elevated linkage disequilibrium (LD) associated with a domestication bottleneck and small effective population sizes (*Ne*). On average, genome-wide heterozygosity of captive Sheppard (0.031) strains shows a 39% reduction compared to wild populations (0.051), supporting the domestication hypothesis (supplementary table 4; supplementary figure 7). Additional captive populations from South America exhibit the highest levels of diversity (0.087) as expected if this represents the ancestral range of the species, followed by African populations. In general, extensively reared captive populations showed lower diversity, with the exception of the European breed of the Sheppard strain which showed levels of diversity similar to that of wild populations. Within-population estimates of LD revealed elevated levels in captive populations consistent with a genetic bottleneck during domestication (supplementary figure 8). This analysis highlighted genome-wide patterns of domestication in captive populations. Next, we investigated which genomic regions showed evidence for adaptation by scanning populations for selective sweeps.

Genetic basis of domestication

Due to strong selection pressures during domestication, mutations carrying a fitness benefit rise in frequency causing linked alleles, by genetic hitchhiking, to sweep with the beneficial allele. Using sweepfinder2 to scan for these signatures of selection we reveal many putative sweeps (n=51) with outlier windows found throughout the genomes of domesticated populations (Figure 2). Selective sweeps were identified firstly at a relaxed threshold when found exceeding the 95th percentile. Numerous sweeps were identified within captive breeds from North America (n=7), Europe (n=40), and the UK, EVE (n=4) belonging to the Sheppard strain (supplementary table 5). Of these potential domestication sweeps three European sweeps overlapped in genomic position with two large North American sweeps, whilst another three overlapped with a single large sweep in the EVE (UK) breed. One sweep was found overlapping across all three Sheppard breeds on chromosome five. A single representative individual from three industry strains (A, B & C), sourced from Asia and

maintained in the UK, were pooled and screened for signatures of selective sweeps to confirm common signatures of selection under artificial captive conditions (supplementary table 5). Several sweeps were identified (n=18) and overlapped with previous candidate domestication loci on chromosome two, four and five (supplementary figure 9). We further explored primary sweeps using a conservative threshold (1%) revealing sweep hotspots on chromosomes two, four and five (supplementary table 5). North American and European Sheppard breeds harboured one and three primary sweeps respectively, whilst EVE contained a single large potential domestication locus manifesting 36.8% (43 Mbp) of chromosome five. Captive industry strains contained two of these primary sweeps, one overlapped with the North American sweep on chromosome two and the second over chromosome five of the EVE strain.

193

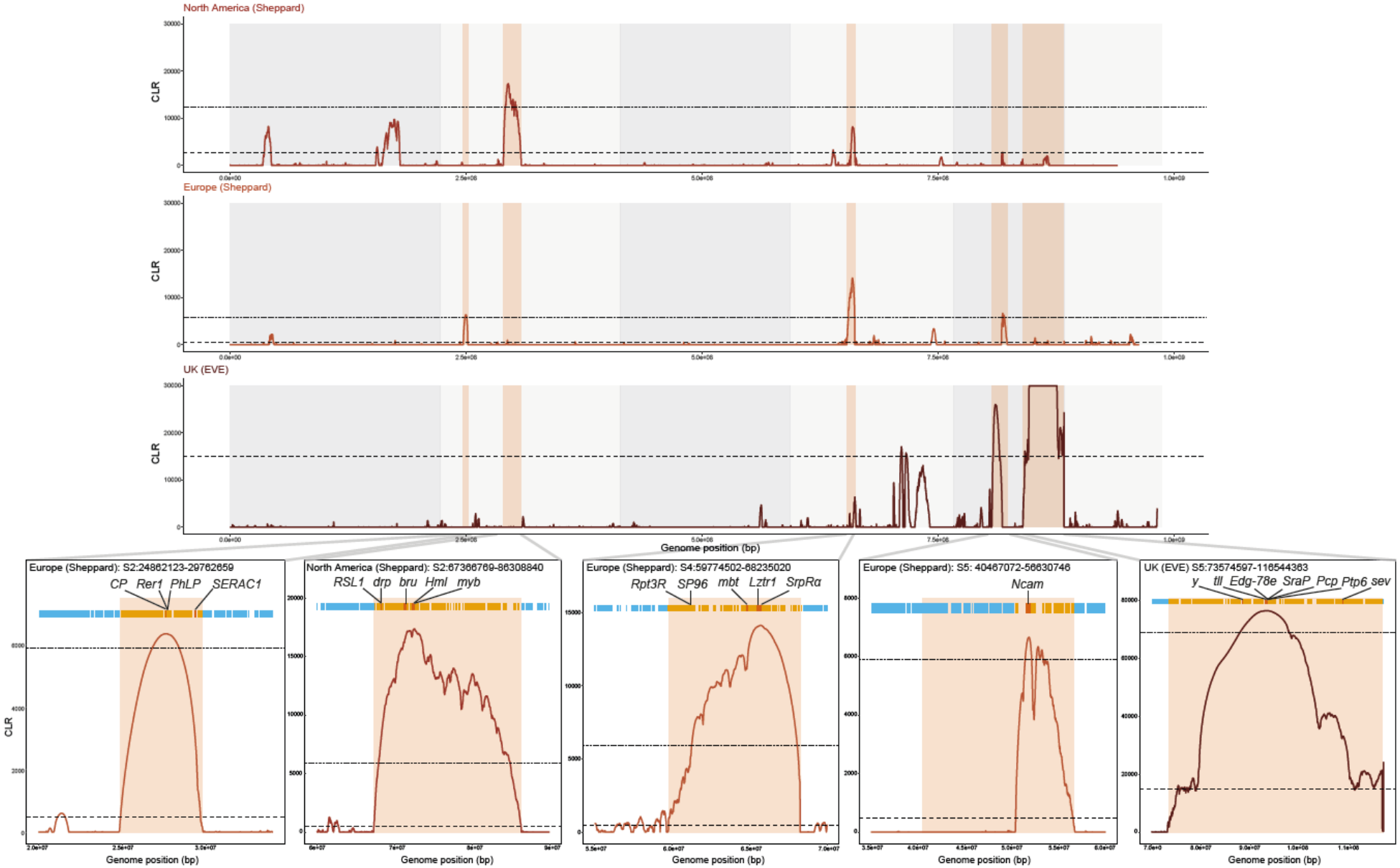

Figure 2. Selective sweeps identified throughout the *Hermetia illucens* genome of domesticated populations inferred from a 1% outlier threshold and magnified within annotated rectangles. Selective sweep Composite Likelihood Ratios (CLR) are capped at 30,000 for data visualisation with 1% (top) and 5% (bottom) thresholds indicated by dashed lines in all populations except UK (EVE) where only the 5% threshold is presented (all thresholds provided in the magnified sweep plot with no CLR caps) due to the 1% threshold exceeding the 30,000 CLR cap. Regions of primary sweeps are enlarged highlighting potential domestication candidate genes found within sweeps and annotated if genes lay within the top 1% window of the sweep or show high F_{ST} or xpEHH support.

We also explored populations recorded as recently established from local wild sources (South America, Africa and Australia) to evaluate whether independent domestication events involve evidence for parallelism over previously identified conservative sweeps. We found evidence for a second independent domestication event which showed convergence in all recent captive populations identifying overlaps with Sheppard strain sweeps at three major loci at the conservative (1%) threshold (supplementary figure 10). At the conservative and relaxed thresholds both South American and African captive populations overlapped with selective sweeps on chromosomes two and four, consistent with sweeps identified in the Sheppard populations (supplementary figure 10). Sweeps within the captive African population also overlapped with the large domestication loci on chromosome five at the relaxed threshold (supplementary figure 10). Australian sweeps overlapping with the North America and European Sheppard breeds were absent but showed multiple overlaps within the large sweep region identified in the EVE breed (S5: 73574597-116544363) (supplementary figure 10).

In order to confirm that these loci are primarily associated with domestication, we also examined wild populations to test whether the five primary domestication loci showed evidence of natural selection in the wild. We found no evidence for selection at candidate domestication loci in the wild on chromosomes two and five, however, the wild North American population harboured a single overlap with the identified domestication loci on chromosome “S4: 59774502–68235020” at the relaxed (5%) threshold (supplementary figure 11). Whereas this analysis supports several of the identified sweeps as potential domestication loci, exclusively occurring in association with domestication of *H. illucens*, sweep “S4: 59774502–68235020” is likely a product of natural selection that provides mutual fitness benefits in an artificial environment or remains an independent signal of domestication due to the large overlapping region identified.

We next used genome-wide scans of genetic differentiation (F_{ST}) and cross population extended haplotype homozygosity (xpEHH) to further scan the genomes for differences between wild and captive Sheppard breeds. Sheppard strains were compared to the closest sampled wild relative, North America, for cross population analysis (supplementary table 6). Peaks of high F_{ST} and xpEHH occurred over previously identified domestication loci on chromosomes two, four and five in all Sheppard breeds (supplementary figure 12 & 13). However, only F_{ST} peaks overlapped with the sweep on chromosome four in the captive North America and European breeds populations (supplementary figure 12). The sweep over chromosome two (S2: 67366769-86308840 bp) harboured elevated F_{ST} of 0.503 (global F_{ST} = 0.122) and xpEHH -$\log_{10}$(*p*-value) of 4.471 (global -$\log_{10}$(*p*-value) = 0.43) values within the North American breeds. The sweep on

chromosome two (S2: 24862123-29762659 bp) in the captive European breed also showed high F_{ST} of 0.631 (global F_{ST} = 0.1) and xpEHH -$\log_{10}$(*p*-value) = 3.266 (global -$\log_{10}$(*p*-value) = 0.428) values. Population specific loci here likely represent different selection pressures in these environments despite a shared history. Both previously identified domestication loci on chromosome five also contained extremely high levels of F_{ST} and xpEHH, particularly over the "S5: 40467072-56630746" region in the North American (max F_{ST} = 0.889; max -$\log_{10}$(*p*-value) = 3.963), European (max F_{ST} = 0.892; max -$\log_{10}$(*p*-value) = 5.369) and EVE (max F_{ST} = 0.822; global F_{ST} = 0.133; max -$\log_{10}$(*p*-value) = 3.823; global -$\log_{10}$(*p*-value) = 0.441) captive breeds. Measures of global F_{ST} can be found in supplementary figure 3. This suggests a common domestication associated mode of genome evolution. Overall, although these signatures are not entirely independent of one another, (non-)overlapping F_{ST} and xpEHH values within selective sweeps further support signatures of domestication selection in these regions.

Hotspots of selection were seen on chromosomes two, four, and five in captive populations (Figure 2). We aimed to identify potential causative candidate genes associated with potential domestication sweeps using the top outlier (1%) regions of each domestication locus as outlined in supplementary table 7. In addition, genes within 20 kb of the top F_{ST} and xpEHH hits within sweeps were identified as candidates. A single 18.94 Mbp domestication peak within the North American captive breed on chromosome S2 (67,366,769-86,308,840 bp) contained 178 genes. Candidates within the outlier region of the sweep had annotated functions involved in egg development (*Myb*) and the innate immune responses (*HMCT*). Supporting hits (xpEHH) identified draper (*drpr*), also activated as part of the immune response, outside of the sweep's outlier region. European breed of the Sheppard strain contained three primary peaks; the chromosome S2 (24,862,123-29,762,659 bp) sweep contained one characterised gene within the outlier region, *SERAC1*, essential for cholesterol trafficking. In addition, two cuticle protein genes and a ciliary regulator (*PhLP*), important for sperm function, were identified within the sweep outside of the outlier region. Two characterised genes were identified within chromosome S4 (59,774,502-68,235,020 bp), including a regulator of (night-time) sleep and brain glycogen metabolism (*Lztr1* or *nowl*), which has been shown to contribute to genetic susceptibility to obesity in fruit flies (Maurer et al. 2020). Two additional genes involved in mushroom body development and photoreceptor differentiation (*mbt*), and axonogensis (*SrpRA*) were identified. The outlier region of the 6.2 Mbp sweep on chromosome S5 (50,409,332-56,630,746 bp) in the European Sheppard breed contained a single gene, neural cell adhesion molecule 2 (*Ncam*), also involved in brain development and sensory perception of smell. The Sheppard genome reference EVE breed showed the strongest selective sweep signal on chromosome S5

(73,274,529-116,544,363 bp), consisting of 662 genes, including ones associated with development, behaviour, and stimuli response. Genes found as top outliers within this sweep included tailless (*tll*), involved in early eye development and male-male aggression as adults. This *tll*-like gene was found clustered with three additional *tll*-like genes. We also found genes involved in pupal development (*Edg-78E* & *Pcp*), the innate immune response (*Sra* & *spz*), and growth (*Ptp61F)*. Outside the outlier region we continue to identify genes associated with behaviour (*y*) and eye development (*sev*). Overall, genes within identified domestication loci are associated with typical traits impacted by domestication primarily affecting development but also including behaviour, metabolism, stimulus responses, and immunity (supplementary figure 14).

Genomic consequences of introgression

Evidence for introgression between the major lineages of *Hermetia illucens* comes from a comparison between nuclear (n=54) and mitochondria (n=112) whole-genome phylogenies, which showed mito-nuclear discordance (Figure 3). We found 12 samples indicating mito-nuclear discordance at the broad level of the major lineages and clades discussed above; all but two of these were localised to mtDNA clade-*B1*. However, overall tree discordance was low affecting only 22% of sampled individuals.

283

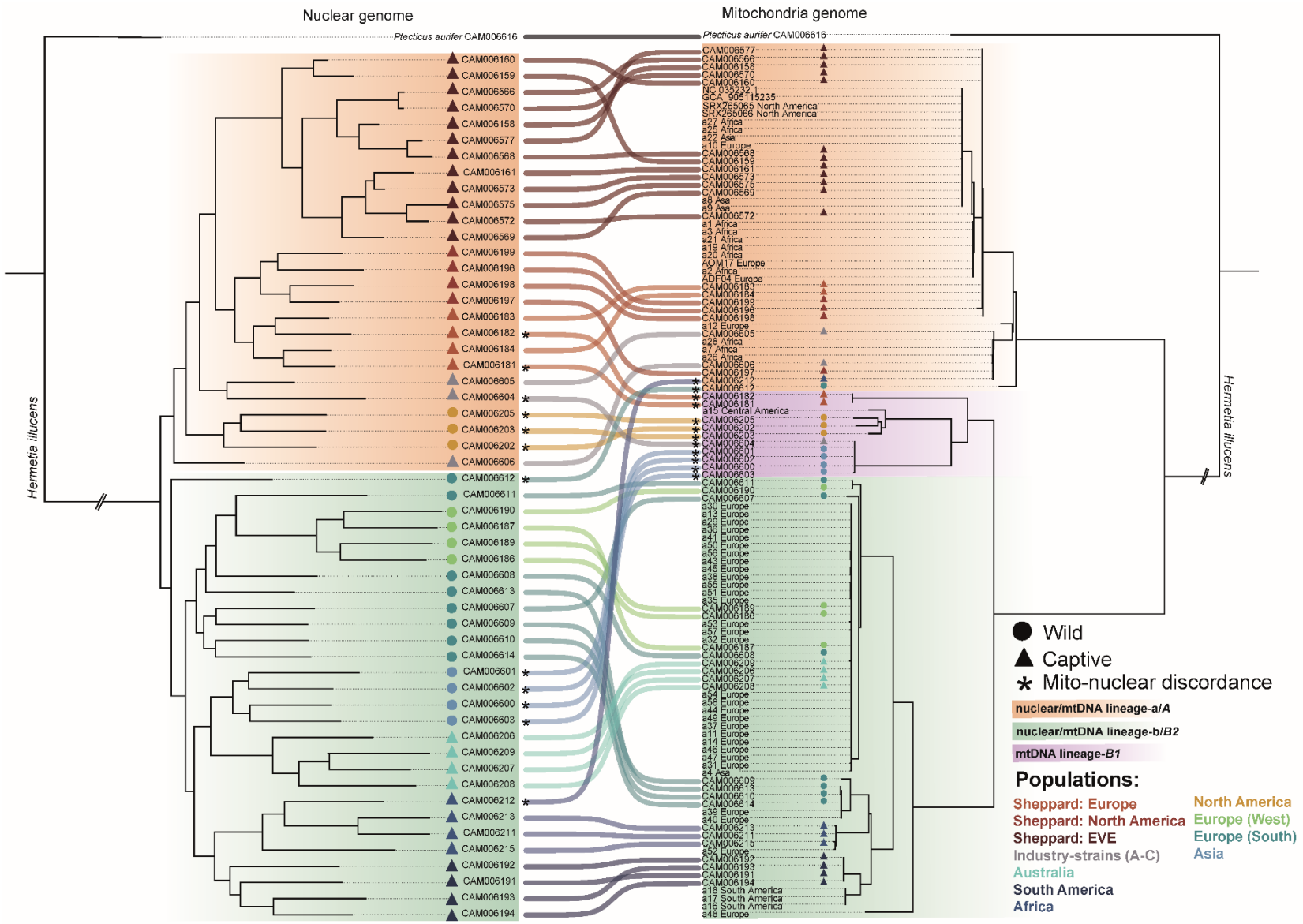

Figure 3. Co-phylogenetic plot of *Hermetia illucens* nuclear (left) and mitochondrial DNA (right) linked using population specific-coloured ribbons and individual CAM ID's. Evidence of mito-nuclear discordance is described by an asterisk in individuals that cross the major identified lineages (*A*, *B1* & *B2*). Whole mitochondria genomes from this study (**I**) and various (**II**) public repositories are used to expand the mtDNA tree.

We next examined genome-wide patterns of introgression within and between divergent lineages. We generated *f*-branch (f_b) statistics using the whole-genome phylogeny of *H. illucens* to estimate genome-wide allele sharing patterns between major lineages using *P. aurifer* as the distant outgroup. Generally, there was little evidence for gene flow between the major lineages, with a few notable exceptions (Figure 4A).

289

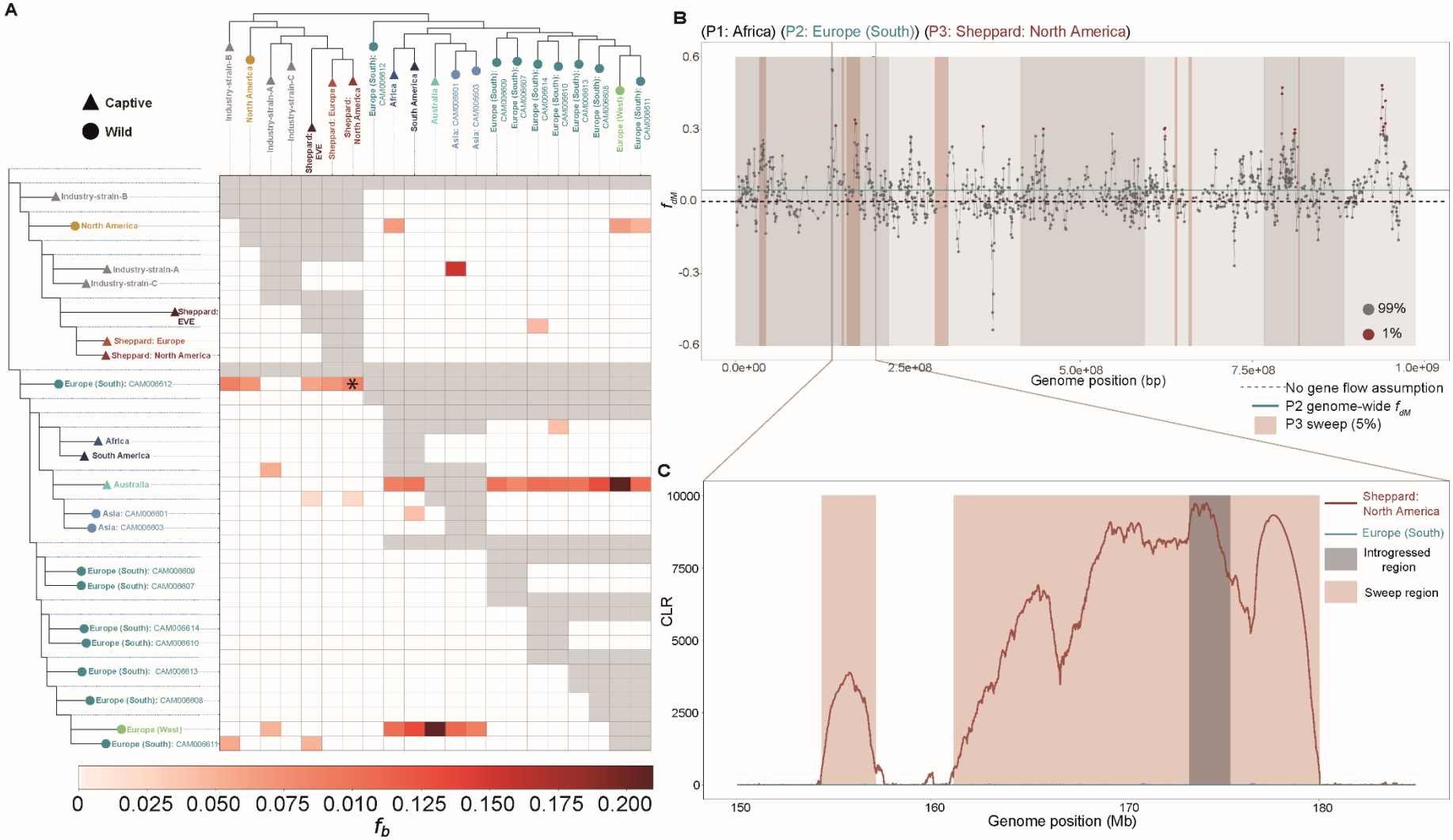

Figure 4. (**A**) Admixture analysis using the *f*-branch (f_b) statistic across sampled *Hermetia illucens* individuals indicating excess allele sharing as indicated by utilised species trees (left and top). Phylogenetic tree on the y-axis represents populations P1 & P2 and is expanded to indicate internal branches enabling comparisons to the P3 population along the x-axis. Admixture is detected between P2 & P3 with P1 fixed as the closest sister group. Grid cell colour and intensity indicates levels of significant allele sharing. Grey shading indicates where the tests cannot be performed due to tree topology. Significant values only are plotted. * Indicates zoomed example for panels B & C. (**B**) Genome-wide f_{DM} plotted highlighting 1% outlier values of allele sharing between P3 and P2. Alternating light and dark grey shading indicate autosomes. (**C**) Plot of selective sweep values indicating adaption in the P3 but not P2 population annotated with sweep and introgression regions.

The strongest evidence of introgression from lineage-*a* into -*b* (f_b = 0.06 - 0.09) occurred from captive lineage-*a* into a wild southern European (CAM00612) representative (Figure 4A). This same individual (CAM00612) was previously identified as showing a mito-nuclear discordant phylogenetic pattern (Figure 3). This may reflect introgression from captive strains escaped into the wild. To identify specific genomic regions implicated, we used f_{dM} to scan the genome revealing an overlapping introgression signal with a selective sweep on chromosome one (Figure 4B). Introgression of genetic material (2.1 Mb) into the wild contained an overrepresentation (93%) of Cytochrome P450s, dramatically increasing gene copy number in this region (supplementary table 8). Whilst the introgressed region from the North American Sheppard breed appears adaptive in the domesticated environment there is no evidence for a selective sweep in the wild southern European populations (Figure 4C). Gene trees group introgressed wild individuals within a Sheppard-derived captive clade, corroborating the hypothesis of introgression (supplementary figure 15). Within lineages, we found evidence for introgression hotspots between lineage-*b*'s wild Europe and captive Australia individuals from multiple global locations, likely a product of anthropogenic mediated dispersal (Figure 4A). We also found evidence for bi-directional gene flow between wild western Europe and Australia indicative of repeated exchange between these geographically distant locations. This evidence supports similar observations of global *H. illucens* colonisation of western Europe via admixed populations including Australasian ancestors (Kaya et al. 2021).

We also detected introgression into an industry strain, with gene flow (f_b = 0.152) into industry-strain-A (Ind-A) from wild Asian populations (Figure 4A). Genome-wide scans of f_{dM} in Ind-A suggests a primarily lineage-*a* genetic background, indicated with a negative genome-wide mean f_{dM}, with a large 7.07 Mb region (S3: 150867943-157941741) of wild Asian origin introgressed on chromosome three (supplementary figure 16), containing a total of 121 genes (supplementary table 9). Several genes across this region have multiple copies including Collagen alpha (*Col4a1*; n=2), Dual oxidase (*Duox*; n=2) and salivary secreted peptide (*Sgsf*; n=5) genes (supplementary table 9). This region did not overlap with previously identified selective sweeps. Such an introgression region provides evidence for possibly adaptive gene flow into captive populations, including those used in industrial settings.

Discussion

Using whole-genome sequencing for major geographic populations of the black soldier fly (*Hermetia illucens*) has revealed the impact of domestication and provided insight into the natural history of this agriculturally important species. Our results describe the genomic footprints of domestication demonstrating evidence for rapid evolution and the impact of converging selection across the genome. We find domesticated populations are largely derived from a North American wild progenitor population and domestication events have occurred independently across the globe. Genetic diversity within this species appears high and potentially harbours a cryptic species complex shown by two divergent nuclear lineages. Despite a worldwide distribution we observe an overall lack of hybridisation between these diverse lineages and observe examples of localised gene flow potentially threatening local wild diversity.

Revealing a putative cryptic species complex

Whole-genome analysis reveals deep structure within *H. illucens*. Previous mtDNA evidence estimated divergence between major clades over 2 my ago (Guilliet et al. 2022), but these data indicate even deeper divergences up to 6.8 my ago, greater than the divergence between *Drosophila species* (Irvin et al. 1998; Caracristi and Schlötterer 2003; Tamura et al. 2004; Verspoor and Haddrill 2011; Nolte et al. 2013; Kao et al. 2015; Kingan et al. 2015; Kang et al. 2019). Combined with a lack of evidence for extensive admixture, our data would suggest that *H. illucens* lineages may perhaps be considered as distinct taxa within a cryptic species complex. However, in the absence of further refined and complete global sampling of underrepresented regions presented in this study this remains speculative. Although previous work, including this study, indicate limited but feasible hybridisation suggesting a

general lack of complete reproductive incompatibilities (Ståhls et al. 2020; Kaya et al. 2021), there is some previous evidence of barriers to random gene flow in *H. illucens* (Hoffmann et al. 2021; Kaya et al. 2021). Nonetheless, further behavioural and morphological analysis would be required before these lineages can be formally described as species.

The existence of cryptic species can clearly have detrimental economic impacts if left undescribed (Bickford et al. 2007; Siddall et al. 2007; Ghai et al. 2014). Progress in breeding programs could be hindered if incompatibilities and/or pre-mating isolation exist between strains (Merot et al. 2017; Maltseva et al. 2021). Cryptic lineages within *H. illucens* are likely in part responsible for the high phenotypic variance previously described (Zhou et al. 2013), including nutritional (Liland et al. 2017; Spranghers et al. 2017; Meneguz et al. 2018) and performance variation (Bosch et al. 2019; Rodrigues et al. 2022). For example, Sandrock et al. (2022) demonstrated significant phenotypic variation among strains of Sheppard and Australasian origin. Within lineage-*a,* different breeds and their (unconscious) interbreeding likely also contributed to this phenotypic variation reported across global studies. Among these samples, lineage-*b* is more diverse and might therefore hold a comparatively larger pool of unrecognised phenotypic variation of possible commercial relevance. Overall, it remains unclear the extent to which commercially motivated worldwide exchange will contribute to the breakdown of these divergent lineages. However, deciphering the evolutionary history of *H. illucens* appears key and timely in order to appropriately discriminate global genetic diversity, as well as improve breeding and targeted genomic selection.

Deciphering the evolutionary history of *H. illucens* provides insights into its biogeographic history. The deep divergences observed (>3 mya) outdate recent expansion into Afro-Eurasia through anthropogenic means (Marshall et al. 2015; Kaya et al. 2021; Guilliet et al. 2022). Therefore, defining evolutionary history for this species can be inferred solely within the Americas. However, in the absence of sufficient data for several key geographic regions to represent a continuous cline throughout the Americas, for example north-eastern South America and the entirety of Central America, divergence by means of isolation-by-distance (IBD) cannot be determined. Based on the data presented in this work several speculative hypotheses regarding the evolutionary history of *H. illucens* can be drawn. Firstly, deep divergence within the Americas occurred during the late Miocene period (~6 mya) and coincided with extreme alpine uplift of the Andean Plateau, potentially driving reproductive barriers between South American populations (Garzione et al. 2006; Ghosh et al. 2006). Evidence of which can be seen in mtDNA patterns (Ståhls et al. 2020; Pazmiño et al. 2023). Secondly, a split between mtDNA lineages-*B1* and -*B2* occurred more recently (~3.2 my) which may represent colonisation of Central America through the bottleneck of

Panama likely driven by means of allopatric speciation. Alternatively, considering the mito-nuclear discordance patterns concentrated in mtDNA lineage-*B1*, particularly southern North American wild populations, there is another possible explanation. Unsampled Central America may represent a potential ancient hybridisation zone where previously isolated indigenous lineage-*a*, or a now extinct ancestral species, made contact with lineage-*b* expanding its ranges northwards from South America. This hypothesis, whilst weak in the absence of Central American samples, may suggest a potential introgression event over 3 million years ago and overlaps with geological records for the formation of the Isthmus of Panama, thus generating a contact zone (O'Dea et al. 2016). Thereby, ancestral lineage-*a* potentially experienced nearly complete nuclear genetic absorbance by lineage-*b*, while still mediating substantial mtDNA introgression (Excoffier et al. 2009). Ancient hybridisation has been shown to facilitate adaptive ecological radiation and speciation (Meier et al. 2017). Likewise, hybridisation may have played an important role in shaping the genetic basis for domestication and consequently, an entire industry largely building on lineage-*a* descendants. These inferences corroborate the presence of cryptic diversity within *H. illucens* and poses major implications on harnessing its breeding potential in an agricultural context. *Hermetia* are distributed throughout the globe yet ~65% of species are found in the Neotropics (Diego and Hauser 2022). However, only *H. illucens* is commonly studied. Therefore, *Hermetia* species pose an interesting system to explore fundamental consequences of hybridisation on geographic and ecological diversification.

Identification of domestication loci with selective sweeps

Black soldier fly domestication has led to a loss of genetic diversity (39%) approaching that of domesticated crops (47-72%) (Bitocchi et al. 2013; Page et al. 2019). This is perhaps surprising considering the recent timescale, estimated around two decades, of *H. illucens* domestication. Continued domestication will likely exacerbate this reduction of genetic diversity in commercial facilities raising further concerns of inbreeding and disease susceptibility (Spielman et al. 2004; Rhode et al. 2020). Our whole-genome nuclear analysis identifies North America as the likely progenitor of domesticated *H. illucens* for the common Sheppard strain. Using COI marker analysis combined with this genome-wide data, we further refine the most likely progenitor of domesticated *H. illucens* to a potential wild North American origin (Ståhls et al. 2020; Kaya et al. 2021; Guilliet et al. 2022). This likely originates to the single domestication event previously documented which established a colony over 200 generations ago (Kaya et al. 2021). In addition to revealing the single domestication event giving rise to the Sheppard strain, this data highlights evidence for at

least a second independent domestication event, maybe more, occurring in captive populations in South America, Africa and Australia.

Unsurprisingly, two decades of domestication occurring in parallel across the globe has resulted in strong but varied selection in the genomes of farmed populations. Some caution should be applied here when interpreting signals stretching over extreme distances of up to >40 Mb, as inbreeding can also lead to runs of homozygosity and signatures similar to selective sweeps. Nonetheless, evidence for shared signatures between descendants of the original and independently domesticated populations lends credence to the hypothesis that these are genuine signatures of selection. The strongest genetic evidence of selection in domesticated populations repeatedly identifies regions containing genes affecting traits commonly associated with domestication such as development, growth and body size (Darwin 1868). Domestication of *H. illucens* has also impacted genes associated with behaviour, reproduction, metabolism, immunity, and various regulatory and maintenance genes. Genes associated with brain development (*mbt, SrpRα* & *Ncam2*) were selected in parallel across the genomes of domesticated *H. illucens*, indicating an impact on brain function in domesticated populations, most likely at the adult stage (Barrett et al. 2022). Genes involved in the development of photoreceptors are also found in parallel across the genome, unsurprising considering light sensitivity has been previously identified as a major determining factor of establishing a successful colony under artificial conditions (Zhang et al. 2010; Oonincx et al. 2016; Heussler et al. 2019; Liu et al. 2020; Macavei et al. 2020). Several genes involved in cuticle protein production were under selection across several life stages potentially indicating factors such as unnatural rearing densities in domesticated conditions may impact commercially relevant traits including body size. This also provides interesting candidates to map the genetic basis of such traits and perform artificial selection and manipulation on body size, previously shown phenotypically to be highly successful (Facchini et al. 2022).

In addition to development, behavioural genes involved in adult male-male aggression may also have been impacted, raising animal welfare concerns if aggressive behaviour is increased (Barrett et al. 2022). Changes in behavioural genes may also have an impact on function. Another gene under selection is known to regulate sleep in humans and fruit flies (*Lztr1* or *nowl*) but also impacts regular metabolic homeostasis maintenance (Maurer et al. 2020). Therefore, it is likely that behavioural patterns, rearing conditions, and nutritional composition are strongly linked in *H. illucens*. Improper photoperiods imposed in artificial environments may therefore impact overall population vigour. Genes involved in circadian rhythm regulation have also been found to be impacted by domestication in both

insect (Xiang et al. 2018) and plant (Müller et al. 2016) species. Overall, these domestication loci are the product of an uncontrolled selection process which may lead to undesirable traits in a commercial context (Rauw et al. 1998).

Evidence of introgression and genomic consequences of gene flow

Overall, there was limited evidence for introgression between major lineages of the *H. illucens* complex. This was unsurprising considering the sampling of major genetics clusters across continents presented here. Despite this sampling method, evidence of hybridisation was identified between lineages occurring in only a small number of individuals in Europe, likely as a product of escaped flies from commercial, or amateur farms as previously suggested by mismatched mitochondrial haplotype patterns of European wild populations (Guilliet et al. 2022). Genetic material found introgressed into a wild southern European representative on chromosome one showed an overrepresentation of P450 genes and had evidence for selection in domesticated populations. This locus perhaps has an adaptive benefit in response to high mycotoxin concentration exposure in industrial organic waste streams (Meijer et al. 2019; Zhan et al. 2019). Whilst a lack of biosecurity measures allowing the release of domesticated flies into the wild poses conservation concerns, it also threatens potential undocumented and economically relevant trait loss from naturally diverse wild populations through genomic homogenisation (McGinnity et al. 2003; Fraser et al. 2010).

There was also some limited evidence for gene flow into active commercial lines belonging to lineage-*a* from wild populations. Wild populations are notoriously difficult to establish in the lab, but might harbour valuable phenotype diversity that could be harnessed by developing introgressed lines through controlled breeding (Kumar et al. 2020; Rhode et al. 2020). Within an industry population we identified an introgressed genomic region derived from a wild Asian source. Whilst no evidence for a selective sweep was found, it would be interesting to further test the fitness consequences of this region. All industry strains (A, B & C) sourced from Asian suppliers showed evidence for admixture between lineage-*a* derived domesticated strains and wild Asian populations. This admixed lineage appears as one of the most prevalent particularly across Asian production facilities (Kaya et al. 2021). Within an industrial setting introgression may promote genetic diversity (Hegstad et al. 2019) and enable the introduction of specific genes for novel applications (Burgarella et al. 2019). Whilst extensive introgression may protect against colony collapse in the short term, establishing an industry with a single founding stock population with a highly similar genetic background severely increases the risk of disease susceptibility in the long term (Bryden et al. 2004; Spielman et al. 2004; Doyle 2016; Rhode et al. 2020). Clearly further work is

needed to understand the degree of heterosis versus hybrid breakdown in crosses between divergent lineages (Lippman and Zamir 2007). In the light of this genetic diversity, cost effective and accessible genetic management will likely play an important role in this industry in the future.

Conclusions

Our study indicates the existence of a previously undescribed cryptic diversity within the species complex of *H. illucens.* Levels of divergence observed are in the same order of magnitude as divergent *Drosophila* species. This suggests a potential need for a taxonomic revaluation of the *H. illucens* complex following further sampling efforts of the Americas. We demonstrate that less than two decades of extensive captive rearing, originating from a wild North American progenitor, has led to dramatic changes in the genome of *H*. *illucens* that show characteristic domestication signatures. Numerous independent and convergent selective sweeps were identified that indicate the consequences of artificial selection. Importantly, irrespective of their global distribution, ongoing domestication of genetically distinct locally naturalised populations show signals of convergent evolution over the same domestication loci. We find domestication of *H. illucens*, through selective sweeps and adaptive introgression, impacts genes associated with development, behaviour, reproduction, metabolism, and immunity. This species is an excellent system to study evolution due to the rapid and ongoing domestication pressures on an industrial and global scale. The genomic resources described here will allow future research to better characterise the genetic identity of strains and to monitor inbreeding in an industrial setting. Elucidating the genetic basis of domestication will make this species an interesting case study for understanding human impacts on domesticated species as well as providing a strain specific foundation for genetic improvements.

Materials and Methods

Sample collection, whole genome sequencing and variant calling

We obtained 54 whole-bodied (adult) *Hermetia illucens* and one *Ptecticus aurifer* sample for whole genome sequencing (WGS). Individuals were obtained from either previously sampled collections (Kaya et al. 2021), maintained (but sourced from three independent Asian suppliers) within industry conditions (Better Origin, UK) or from a taxonomic voucher set (source: M. Hauser). Total genomic DNA was extracted from thorax muscle tissue and purified using the Qiagen Blood & Culture DNA Midi Kit (Qiagen, Germany). Libraries (*n* = 39) were prepared, and genome sequenced by BGI (BGI, Hong Kong) as in Generalovic et

al. (2021), including the 12 previously published genomes. Novogene (Cambridge, UK) constructed a second set of libraries (*n* = 11) and performed paired end (PE) 150 bp sequencing on the NovaSeq platform. The remaining five libraries were sequenced in the same manner but constructed using a custom Tn5 based protocol (Picelli et al. 2014). Raw sequencing data was quality checked, with FastQC v0.11.4 (Andrews 2010) and visualised in MultiQC v1.9 (Ewels et al. 2016) both pre- and post- adapter trimming with cutadapt v1.8.1 (Martin 2011). Treated reads were mapped to the GCA_905115235.1 reference genome (Generalovic et al. 2021) using bwa-mem v0.7.17 algorithm (Li and Durbin 2009). Resulting BAM files were sorted, duplicates marked and removed using samtools v1.9 (Li et al. 2009) and picard v2.9.2 (http://broadinstitute.github.io/picard). Variant calling was performed following the best practices of Genome Analysis Toolkit (GATK) v3.7.0 (McKenna et al. 2010). Genotyped SNPs were quality checked and filtered for high quality (>75) sites setting a minimum and maximum depth cut-off (10-34X) and ignoring sites containing a high proportion of missingness (>15%). Assessment of contaminants based on allelic balance within samples was carried out using allelicBalance.py (https://github.com/joanam/scripts/blob/master/allelicBalance.py). An additional VCF dataset was generated filtering for minor allele frequencies using a maf threshold of 0.1.

Mitochondrial *de novo* assembly

Analysis of mitochondrial genomes was performed by assembling *de novo* whole mitochondrial genomes using a subset of the novel WGS reads produced in this study and collected from previous work (Qi et al. 2017; Generalovic et al. 2021; Guilliet et al. 2022). We assumed, due to high mitochondria copy number in the muscle tissue used for DNA extraction, that mitochondrial DNA was enriched in our WGS data. We used a minimum subset of 1.6 Gb raw PE data as input for *de novo* assembly per individual and a maximum of 4.8 Gb (supplementary table 10). Mitochondrial *de novo* assembly was carried out using the MitoZ toolkit v2.4 (Meng et al. 2019). We used the ‘all’ function to assemble, annotate and plot genomes with parameters ‘genetic_code=5’ and ‘clade=Arthropoda’. In cases where assemblies yielded the reverse compliment of the mitogenome, identified by the gene order and orientation being the opposite to that of the ancestral insect (Cameron 2014), the mitogenome was derived by running the assembly through the reverse compliment function of MEGA-X v10.2.6 (Kumar et al. 2018). An additional 57 mitochondrial genomes were obtained from previous studies (Qi et al. 2017; Generalovic et al. 2021; Guilliet et al. 2022).

Phylogenetic and population structure

A Maximum Likelihood (ML) phylogenetic tree was constructed using whole nuclear genomes to assess phylogenetic structure and relationships within the *H. illucens* species. We used RAxML program v8.2.9 (Stamatakis 2014) to construct a ML tree using the generalized time-reversible substitution model and gamma (GTRGAMMA) model, Lewis ascertainment bias and 100 bootstrap (BS) replications based on a random subset of 648,059 nuclear whole-genome SNPs. Further phylogenetic analysis using ASTRAL v5.7.8 (Zhang et al. 2018), a coalescent based method robust to Incomplete Lineage Sorting (ILS), was performed to validate the ML analysis. A 10% sample of SNPs from each chromosome was extracted and split into 5,000 SNP segments and a gene tree was constructed on each using 100 BS replications totalling 8,800 trees, this was repeated for both the Neighbour-Joining (NJ) and ML method.

Phylogenetic reconstruction of the mitochondrial genomes was also performed. Whole mitochondrial genomes were aligned using MUSCLE v3.8.31 (Edgar 2004) and a phylogenetic ML tree produced using RAxML v8.2.9 (Stamatakis 2014). A principal component analysis (PCA) (Patterson et al. 2006) was performed using plink v1.9 (Purcell et al. 2007) with *H. illucens* samples only. A random subsample of two individuals per unique geographic sampled population was used to conduct the PCA analysis after it was found that an unbalanced design artificially inflated genetic variation across both PC1 and PC2. Structure analysis was performed using ADMIXTURE v1.3.0 (Alexander et al. 2009) assuming a minimum and maximum number of ancestral populations (*k*) of two and 11, respectively, based on the number of populations sampled. Linkage pruning was performed on SNPs used for both PCA and STRUCTURE analysis using plink v1.9 (Purcell et al. 2007). A NJ tree (Saitou and Nei 1987) was constructed using wild and captive Cytochrome Oxidase I (COI) mitochondrial marker genes from published and *de novo* assembled mitochondrial genomes. A NJ analysis of COI sequences was performed by aligning sequences with MUSCLE v3.8.31 (Edgar 2004) and phylogenetic analysis with MEGA v10.1.7 (Kumar et al. 2018) using default parameters.

Intraspecific divergence and genetic differentiation

Patterns of intraspecific divergence within the *H. illucens* complex were revealed using genome wide (nuclear) scans for genetic differentiation (F_{ST}) and sequence divergence (D_{XY}) on all populations. Non-overlapping 20 kb windows in 20 kb steps containing a minimum of 100 genotyped sites (-w 20000 -m 100 -s 20000) were used to generate global genome-wide F_{ST} and D_{XY} statistics using popgenWindows.py (https://github.com/simonhmartin/genomics_general).

Divergence time analysis

Due to the paucity of *H. illucens* divergence data, a dated Diptera tree was required to provide temporal calibration points for a *H. illucens* divergence analysis, as in Nardi *et al.* (2010). Here, we used mitochondrial DNA (mtDNA) only to estimate approximate divergences times. Mitogenomes of two *H. illucens* individuals from divergent branches of both the nuclear and mitochondrial *H. illucens* phylogeny and a single outgroup individual, *Ptecticus aurifer* were obtained along with ten Diptera mitogenomes from GenBank (supplementary table 11) (Agarwala et al. 2018). All Diptera mitochondrial genomes were aligned in MUSCLE v3.8.31 (Edgar 2004). A molecular clock test was conducted on the resultant alignment and the optimal model of sequence evolution was selected by AIC using MEGA-X v10.2.6 (Kumar et al. 2018). A Bayesian dating analysis was conducted in BEAST v2.6.6 (Bouckaert et al. 2019), using a GTR+G+I model of sequence evolution, and a relaxed lognormal molecular clock. Dating priors were set to calibrate six nodes spanning the Diptera phylogeny and the Yule model was used for the tree prior (supplementary table 12). Markov chain Monte Carlo (MCMC) length was set to 100 million, with samples logged every 1,000 steps. The model output was inspected in Tracer v1.7.2 (Rambaut et al. 2018). After checking for parameter convergence, the burn-in was set to 25%, and the effective sample size of relevant parameters was verified to exceed 150. The dated phylogeny was derived using TreeAnnotator package of BEAST v2.6.6 (Bouckaert et al. 2019). To get estimates for key divergences within the *H. illucens* phylogeny, individuals which spanned the phylogeny were used (supplementary table 13), alongside a *P. aurifer* outgroup. Alignment and divergence analysis was repeated a second time using the parameters previously described. We used a tree prior with a coalescent model of constant population size. Temporal constraints were placed on two nodes, using the output of the initial BEAST run, corresponding to the *P. aurifer* - *H. illucens* split and the common ancestor of the *H. illucens* clade (supplementary table 12). The MCMC chain was as previously described. After inspecting the model output in Tracer v1.7.2 (Rambaut et al. 2018), a burn-in of 10% was set.

Genetic signals of domestication

Genomic diversity, and Linkage Disequilibrium (LD) decay were calculated to reveal genetic signatures of domestication in sampled populations with a sequencing depth >19x. Nucleotide diversity (π) was calculated in non-overlapping 100 kb windows using popgenWindows.py (-w 100000 -m 100 -s 100000) (https://github.com/simonhmartin/genomics_general) using 100 kb steps containing at least 100 genotyped sites. Values of π for every individual was presented as mean heterozygosity

within each sampled population. Patterns of LD decay were first calculated genome wide for a random subset of three individuals for each population. We used PopLDdecay v3.41 (Zhang et al. 2019) to generate pairwise LD r^2 values setting a maximum distance of 50 kb between sampled SNPs.

Identifying genetic basis of domestication

To dissect the genetic basis of domestication in *H. illucens* we performed scans for selective sweeps to reveal genomic regions associated with positive selection. We tested each maintained captive population independently despite known shared population history and strong *k*-means clustering support for a single population structure of the Sheppard-lineage, due to a documented split approximately 1.5 decades ago. We also used patterns observed from *k*-means ancestry assignment (admixture) analysis to define individuals into groups for selective sweep detection. All geographic populations were treated independently based off results at $k = 11$. Industry strains (A-C) were grouped based on population structure results and identical rearing conditions. We used a composite likelihood approach to detect beneficial mutations that have rapidly increased in frequency causing a selective sweep using SweepFinder2 (Degiorgio et al. 2016). A genome wide empirical frequency spectrum was generated for each tested population and sweep detection was performed using sites spaced 20 kb apart (-lg 20000). All selective sweep tests were set an arbitrary relaxed outlier threshold of 5% to identify sweeps associated with selection followed by a 1% threshold to conservatively identify primary sweeps associated with selection. Sweep co-ordinates were quantified by CLR values exceeding the genome-wide average for each respective population or if the flanking region of sweeps occurred within 50 kb of another sweep, in which case they were treated as a single sweep. These key sweeps were compared to populations in ongoing domestication and the wild to confirm association with domestication. Additional support using F_{ST} and comparative cross population haplotype statistics (xpEHH) with the closest sampled wild progenitor was acquired. Population specific F_{ST} was calculated genome-wide as previously described. Cross Population Extended Haplotype Homozygosity (xpEHH) was generated using the rehh v3.2.1 package (Gautier et al. 2017) in R v3.6.3 (RStudio Team 2019). We also used a 1% threshold to identify causative candidate genes within key selective sweeps associated with selection. Candidate loci were identified from gene codes in the GCA_905115235.1 reference by performing blastp searches against the Arthropoda blast nr database using an *e*-value threshold of 1*e*-5. Gene Ontology summary information for each candidate gene was obtained from flybase (Thurmond et al. 2019).

Introgression and gene flow

Mito-nuclear discordance was predicted by plotting a co-phylogenetic plot of the previously constructed nuclear and mitochondrial phylogenies using the phytools v1.0-3 package (Revell 2012) in R v3.6.3 (RStudio Team 2019). Patterns of admixture were detected by revealing excess allele sharing between individuals using ancestral (ABBA) and derived (BABA) allele patterns (Green et al. 2010; Durand et al. 2011), f_{dM} values (Malinsky et al. 2015) and the *f*-branch statistic (Malinsky et al. 2018). We used ABBA/BABA patterns using Dsuite v0.4r38 Dtrios function (Malinsky et al. 2021) and provided the previously resolved SNP based phylogeny (Figure 1A) as a tree file (-t). Trios showing *p*-values > 0.05 and Z-scores < 3 post *p*-value adjustment using Benjamini-Hochberg multiple correction in R v4.0.3 (RStudio Team 2019) were filtered out of the analysis. We generated the *f*-branch (f_b) statistic using individuals sequenced to the highest coverage to represent populations within the phylogeny to assign gene flow to all, including internal ancestral, branches using *P. aurifer* as the outgroup (Malinsky et al. 2018). Populations that were not monophyletic were represented by multiple individuals in the f_b analysis. Genome-wide scans were performed on samples showing an excess of gene flow identified from the *f*-branch method using the Dsuite Dinvestigate function calculating average f_{dM} over 50 SNPs in 25 SNP steps across the whole genome. Genomic hotspots of introgression were identified by setting an arbitrary 1% cut-off for f_{dM} values.

Acknowledgments

This work was supported by the Biotechnology and Biological Sciences Research Council (BB/M011194/1).

Availability of supporting data

The data sets supporting the results of this article are available under Bioprojects PRJEB37575 and PRJEB58720.

References

Agarwala R, Barrett T, Beck J, Benson DA, Bollin C, Bolton E, Bourexis D, Brister JR, Bryant SH, Canese K, et al. 2018. Database resources of the National Center for Biotechnology Information. *Nucleic Acids Res.* 46:D8–D13.

Alexander DH, Novembre J, Lange K. 2009. Fast model-based estimation of ancestry in unrelated individuals. *Genome Res.* 19:1655–1664.

Andrews S. 2010. FastQC: a quality control tool for high throughput sequence data.

Barrett M., Chia SY, Fischer B, Tomberlin JK. 2022. Welfare considerations for farming black soldier flies, *Hermetia illucens* (Diptera: Stratiomyidae): a model for the insects as food and feed industry. *J. Insects as Food Feed*:1–30.

Barrett M, Godfrey KR, Sterner EJ, Waddell EA. 2022. Impacts of development and adult sex on brain cell numbers in the Black Soldier Fly, *Hermetia illucens* L. (Diptera: Stratiomyidae). *Arthropod Struct. Dev.* 70:101174.

Bickford D, Lohman DJ, Sodhi NS, Ng PKL, Meier R, Winker K, Ingram KK, Das I. 2007. Cryptic species as a window on diversity and conservation. *Trends Ecol. Evol.* 22:148–155.

Bitocchi E, Bellucci E, Giardini A, Rau D, Rodriguez M, Biagetti E, Santilocchi R, Spagnoletti Zeuli P, Gioia T, Logozzo G, et al. 2013. Molecular analysis of the parallel domestication of the common bean (*Phaseolus vulgaris*) in Mesoamerica and the Andes. *New Phytol.* 197:300–313.

Bosch G, van Zanten HHE, Zamprogna A, Veenenbos M, Meijer NP, van der Fels-Klerx HJ, van Loon JJA. 2019. Conversion of organic resources by black soldier fly larvae: Legislation, efficiency and environmental impact. *J. Clean. Prod.* 222:355–363.

Bouckaert R, Vaughan TG, Barido-Sottani J, Duchêne S, Fourment M, Gavryushkina A, Heled J, Jones G, Kühnert D, De Maio N, et al. 2019. BEAST 2.5: An advanced software platform for Bayesian evolutionary analysis. *PLoS Comput. Biol.* 15:1–28.

BroadInstitute. Picard http://broadinstitute.github.io/picard.

Bryden CA, Heath JW, Heath DD. 2004. Performance and heterosis in farmed and wild Chinook salmon (*Oncorhynchus tshawytscha*) hybrid and purebred crosses. *Aquaculture* 235:249–261.

Burgarella C, Barnaud A, Kane NA, Jankowski F, Scarcelli N, Billot C, Vigouroux Y, Berthouly-Salazar C. 2019. Adaptive introgression: An untapped evolutionary mechanism for crop adaptation. *Front. Plant Sci.* 10:1–17.

Cameron SL. 2014. Insect mitochondrial genomics: Implications for evolution and phylogeny. *Annu. Rev. Entomol.* 59:95–117.

Caracristi G, Schlötterer C. 2003. Genetic differentiation between American and European *Drosophila melanogaster* populations could be attributed to admixture of African alleles. *Mol. Biol. Evol.* 20:792–799.

Darwin C. 1868. The variation of animals and plants under domestication. J. murray

Degiorgio M, Huber CD, Hubisz MJ, Hellmann I, Nielsen R. 2016. SweepFinder2: Increased sensitivity, robustness and flexibility. *Bioinformatics* 32:1895–1897.

Diego FA, Hauser M. 2022. Large Flies Overlooked : the Genus *Hermetia Latreille*, 1804 (Diptera : Stratiomyidae) in the Neotropics , with 11 Synonyms and a New Species to Brazil. *Neotrop. Entomol.* 1804.

Doyle RW. 2016. Inbreeding and disease in tropical shrimp aquaculture: A reappraisal and caution. *Aquac. Res.* 47:21–35.

Durand EY, Patterson N, Reich D, Slatkin M. 2011. Testing for ancient admixture between closely related populations. *Mol. Biol. Evol.* 28:2239–2252.

Edgar RC. 2004. MUSCLE: A multiple sequence alignment method with reduced time and space complexity. *BMC Bioinformatics* 5:1–19.

Ewels P, Magnusson M, Lundin S, Kaller M. 2016. MultiQC: summarize analysis results for multiple tools and samples in a single report. *Bioinformatics* 32:3047–3048.

Excoffier L, Foll M, Petit RJ. 2009. Genetic consequences of range expansions. *Annu. Rev. Ecol. Evol. Syst.* 40:481–501.

Facchini E, Shrestha K, van den Boer E, Junes P, Sader G, Peeters K, Schmitt E. 2022. Long-Term Artificial Selection for Increased Larval Body Weight of *Hermetia illucens* in Industrial Settings. *Front. Genet.* 13:1–9.

Fraser DJ, Minto C, Calvert AM, Eddington JD, Hutchings JA. 2010. Potential for domesticated-wild interbreeding to induce maladaptive phenology across multiple populations of wild Atlantic salmon (*Salmo salar*). *Can. J. Fish. Aquat. Sci.* 67:1768–1775.

Garzione CN, Molnar P, Libarkin JC, MacFadden BJ. 2006. Rapid late Miocene rise of the Bolivian Altiplano: Evidence for removal of mantle lithosphere. *Earth Planet. Sci. Lett.* 241:543–556.

Gautier M, Klassmann A, Vitalis R. 2017. rehh 2.0: a reimplementation of the R package rehh to detect positive selection from haplotype structure. *Mol. Ecol. Resour.* 17:78–90.

Generalovic TN, McCarthy SA, Warren IA, Wood JMD, Torrance J, Sims Y, Quail M, Howe K, Pipan M, Durbin R, et al. 2021. A high-quality, chromosome-level genome assembly of the Black Soldier Fly (*Hermetia illucens* L.). *G3 Genes, Genomes, Genet.* 11.

Ghai RR, Chapman CA, Omeja PA, Davies TJ, Goldberg TL. 2014. Nodule Worm Infection in Humans and Wild Primates in Uganda: Cryptic Species in a Newly Identified Region of Human Transmission. *PLoS Negl. Trop. Dis.* 8:39.

Ghosh P, Garzione CN, Eiler JM. 2006. Rapid Uplift of the Altiplano Revealed Through 13C-18O Bonds in Paleosol Carbonates. *Science.* 311:511–515.

Gold M, Tomberlin JK, Diener S, Zurbrügg C, Mathys A. 2018. Decomposition of biowaste macronutrients, microbes, and chemicals in black soldier fly larval treatment: A review. *Waste Manag.* 82:302–318.

Green RE, Krause J, Briggs AW, Maricic T, Stenzel U, Kircher M, Patterson N, Li H, Zhai W, Fritz MHY, et al. 2010. A draft sequence of the neandertal genome. *Science.* 328:710–722.

Guilliet J, Baudouin G, Pollet N, Filée J. 2022. What complete mitochondrial genomes tell us about the evolutionary history of the black soldier fly, *Hermetia illucens*. *BMC Ecol. Evol.*:1–15.

Hegstad JM, Nelson RL, Renny-Byfield S, Feng L, Chaky JM. 2019. Introgression of novel genetic diversity to improve soybean yield. *Theor. Appl. Genet.* 132:2541–2552.

Heussler CD, Walter A, Oberkofler H, Insam H, Arthofer W, Schlick-Steiner BC, Steiner FM. 2019. Correction: Influence of three artificial light sources on oviposition and half-life of the Black Soldier Fly, *Hermetia illucens* (Diptera: Stratiomyidae): Improving small-scale indoor rearing (PLoS ONE (2018) 13:5 (e0197896) DOI: 10.1371/journal.pone.0197. *PLoS One* 14:40–45.

Hoffmann L, Hull KL, Bierman A, Badenhorst R, Bester-van der Merwe AE, Rhode C. 2021. Patterns of Genetic Diversity and Mating Systems in a Mass-Reared Black Soldier Fly Colony. *Insects* 12:480.

Van Huis A. 2013. Potential of insects as food and feed in assuring food security. *Annu. Rev. Entomol.* 58:563–583.

Van Huis A, Van Itterbeeck J, Klunder H et al. 2013. Edible Insects: Future Prospects for Food and Feed Security.

Irvin SD, Wetterstrand KA, Hutter CM, Aquadro CF. 1998. Genetic variation and differentiation at microsatellite loci in *Drosophila simulans*: Evidence for founder effects in new world populations. *Genetics* 150:777–790.

Kang L, Rashkovetsky E, Michalak K, Garner HR, Mahaney JE, Rzigalinski BA, Korol A, Nevo E, Michalak P. 2019. Genomic divergence and adaptive convergence in *Drosophila simulans* from Evolution Canyon, Israel. *Proc. Natl. Acad. Sci. U. S. A.* 116:11839–11844.

Kao JY, Zubair A, Salomon MP, Nuzhdin S V., Campo D. 2015. Population genomic analysis uncovers African and European admixture in *Drosophila melanogaster* populations from

the south-eastern United States and Caribbean Islands. *Mol. Ecol.* 24:1499–1509.

Kareiva P, Watts S, McDonald R, Boucher T. 2007. Domesticated nature: Shaping landscapes and ecosystems for human welfare. *Science.* 316:1866–1869.

Kaya C, Generalovic TN, Ståhls G, Hauser M, Samayoa AC, Nunes-Silva CG, Roxburgh H, Wohlfahrt J, Ewusie EA, Kenis M, et al. 2021. Global population genetic structure and demographic trajectories of the black soldier fly, *Hermetia illucens*. *BMC Biol.* 19:1–22.

Kingan SB, Geneva AJ, Vedanayagam JP, Garrigan D, Garrigan D. 2015. Genome divergence and gene flow between *Drosophila simulans* and *D. mauritiana.*

Kumar A, Sandhu N, Venkateshwarlu C, Priyadarshi R, Yadav S, Majumder RR, Singh VK. 2020. Development of introgression lines in high yielding, semi-dwarf genetic backgrounds to enable improvement of modern rice varieties for tolerance to multiple abiotic stresses free from undesirable linkage drag. *Sci. Rep.* 10:1–13.

Kumar S, Stecher G, Li M, Knyaz C, Tamura K. 2018. MEGA X: Molecular Evolutionary Genetics Analysis across Computing Platforms.Battistuzzi FU, editor. *Mol. Biol. Evol.* 35:1547–1549.

Kwon YM, Vranken N, Hoge C, Lichak MR, Norovich AL, Francis KX, Camacho-Garcia J, Bista I, Wood J, McCarthy S, et al. 2022. Genomic consequences of domestication of the Siamese fighting fish. *Sci. Adv.* 8:1–17.

Larson G, Piperno DR, Allaby RG, Purugganan MD, Andersson L, Arroyo-Kalin M, Barton L, Vigueira CC, Denham T, Dobney K, et al. 2014. Current perspectives and the future of domestication studies. *Proc. Natl. Acad. Sci. U. S. A.* 111:6139–6146.

Lecocq T, Lecocq T, The I, Domestication D, Fabrice H, Animal T, Lecocq T. 2019. Insects: The Disregarded Domestication Histories.

Li H, Durbin R. 2009. Fast and accurate short read alignment with Burrows-Wheeler transform. *Bioinformatics* 25:1754–1760.

Li H, Handsaker B, Wysoker A, Fennell T, Ruan J, Homer N, Marth G, Abecasis G, Durbin R. 2009. The Sequence Alignment/Map format and SAMtools. *Bioinformatics* 25:2078–2079.

Liland NS, Biancarosa I, Araujo P, Biemans D, Bruckner CG, Waagbø R, Torstensen BE, Lock EJ. 2017. Modulation of nutrient composition of black soldier fly (*Hermetia illucens*) larvae by feeding seaweed-enriched media. *PLoS One* 12:1–23.

Lippman ZB, Zamir D. 2007. Heterosis: revisiting the magic. *Trends Genet.* 23:60–66.

Liu Z, Najar-Rodriguez AJ, Minor MA, Hedderley DI, Morel PCH. 2020. Mating success of

the black soldier fly, *Hermetia illucens* (Diptera: Stratiomyidae), under four artificial light sources. *J. Photochem. Photobiol. B Biol.* 205:1–8.

Mabry ME, Turner-Hissong SD, Gallagher EY, McAlvay AC, An H, Edger PP, Moore JD, Pink DAC, Teakle GR, Stevens CJ, et al. 2021. The Evolutionary History of Wild, Domesticated, and Feral *Brassica oleracea* (Brassicaceae). *Mol. Biol. Evol.* 38:4419–4434.

Macavei LI, Benassi G, Stoian V, Maistrello L. 2020. Optimization of *Hermetia illucens* (L.) egg laying under different nutrition and light conditions. *PLoS One* 15:1–18.

Malinsky M, Challis RJ, Tyers AM, Schiffels S, Terai Y, Ngatunga BP, Miska EA, Durbin R, Genner MJ, Turner GF. 2015. Genomic islands of speciation separate cichlid ecomorphs in an East African crater lake. *Science.* 350:1493–1498.

Malinsky M, Matschiner M, Svardal H. 2021. Dsuite - Fast D-statistics and related admixture evidence from VCF files. *Mol. Ecol. Resour.* 21:584–595.

Malinsky M, Svardal H, Tyers AM, Miska EA, Genner MJ, Turner GF, Durbin R. 2018. Whole-genome sequences of Malawi cichlids reveal multiple radiations interconnected by gene flow. *Nat. Ecol. Evol.* 2:1940–1955.

Maltseva AL, Varfolomeeva MA, Lobov AA, Tikanova PO, Repkin EA, Babkina IY, Panova M, Mikhailova NA, Granovitch AI. 2021. Premating barriers in young sympatric snail species. 1–16.

Marshall SA, Woodley NE, Hauser M. 2015. The historical spread of the Black Soldier Fly, *Hermetia illucens* (L.) (Diptera, Stratiomyidae, Hermetiinae), and its establishment in Canada. *J. Entomol. Soc. Ontario* 146:51–54.

Martin M. 2011. Cutadapt removes adapter sequences from high-throughput sequencing reads. *EMBnet.journal* 17:10.

Martin S. Genomics General. Available from: https://github.com/simonhmartin/genomics_general

Maurer GW, Malita A, Nagy S, Koyama T, Werge TM, Halberg KA, Texada MJ, Rewitz K. 2020. Analysis of genes within the schizophrenia-linked 22q11.2 deletion identifies interaction of night owl/LZTR1 and NF1 in GABAergic sleep control.

McGinnity P, Prodöhl P, Ferguson A, Hynes R, Maoiléidigh NÓ, Baker N, Cotter D, O'Hea B, Cooke D, Rogan G, et al. 2003. Fitness reduction and potential extinction of wild populations of Atlantic salmon, Salmo salar, as a result of interactions with escaped farm salmon. *Proc. R. Soc. B Biol. Sci.* 270:2443–2450.

McKenna A, Hanna M, Banks E, Sivachenko A, Cibulskis K, Kernytsky A, Garimella K, Altshuler D, Gabriel S, Daly M, et al. 2010. The Genome Analysis Toolkit: A MapReduce framework for analyzing next-generation DNA sequencing data. *Genome Res.* 20:1297–1303.

Meier JI, Marques DA, Mwaiko S, Wagner CE, Excoffier L, Seehausen O. 2017. Ancient hybridization fuels rapid cichlid fish adaptive radiations. *Nat. Commun.* 8:1–11.

Meijer N, Stoopen G, van der Fels-Klerx HJ, van Loon JJA, Carney J, Bosch G. 2019. Aflatoxin B1 Conversion by Black Soldier Fly (*Hermetia illucens*) Larval Enzyme Extracts. *Toxins (Basel).* 11:532.

Meneguz M, Schiavone A, Gai F, Dama A, Lussiana C, Renna M, Gasco L. 2018. Effect of rearing substrate on growth performance, waste reduction efficiency and chemical composition of black soldier fly (*Hermetia illucens*) larvae. *J. Sci. Food Agric.* 98:5776–5784.

Meng G, Li Y, Yang C, Liu S. 2019. MitoZ: A toolkit for animal mitochondrial genome assembly, annotation and visualization. *Nucleic Acids Res.* 47.

Merot C, Salazar C, Merrill RM, Jiggins CD, Joron M. 2017. What shapes the continuum of reproductive isolation ? Lessons from Heliconius butterflies.

Müller NA, Wijnen CL, Srinivasan A, Ryngajllo M, Ofner I, Lin T, Ranjan A, West D, Maloof JN, Sinha NR, et al. 2016. Domestication selected for deceleration of the circadian clock in cultivated tomato. *Nat. Publ. Gr.* 48.

Nardi F, Carapelli A, Boore JL, Roderick GK, Dallai R, Frati F. 2010. Domestication of olive fly through a multi-regional host shift to cultivated olives: Comparative dating using complete mitochondrial genomes. *Mol. Phylogenet. Evol.* 57:678–686.

Nolte V, Pandey RV, Kofler R, Schlötterer C. 2013. Genome-wide patterns of natural variation reveal strong selective sweeps and ongoing genomic conflict in Drosophila mauritiana. *Genome Res.* 23:99–110.

O'Dea A, Lessios HA, Coates AG, Eytan RI, Restrepo-Moreno SA, Cione AL, Collins LS, De Queiroz A, Farris DW, Norris RD, et al. 2016. Formation of the Isthmus of Panama. *Sci. Adv.* 2:1–12.

Oonincx DGAB, Volk N, Diehl JJE, van Loon JJA, Belušič G. 2016. Photoreceptor spectral sensitivity of the compound eyes of black soldier fly (*Hermetia illucens*) informing the design of LED-based illumination to enhance indoor reproduction. *J. Insect Physiol.* 95:133–139.

Ostrander EA, Wayne RK, Freedman AH, Davis BW. 2017. Demographic history, selection

and functional diversity of the canine genome. *Nat. Rev. Genet.* 18:705–720.

Page A, Gibson J, Meyer RS, Chapman MA. 2019. Eggplant domestication: Pervasive gene flow, feralization, and transcriptomic divergence. *Mol. Biol. Evol.* 36:1359–1372.

Patterson N, Price AL, Reich D. 2006. Population structure and eigenanalysis. *PLoS Genet.* 2:2074–2093.

Pazmiño MF, Hierro AG Del, Flores FJ. 2023. Genetic diversity and organic waste degrading capacity of *Hermetia illucens* from the evergreen forest of the Equatorial Choco lowland. 1–21.

Picelli S, Björklund ÅK, Reinius B, Sagasser S, Winberg G, Sandberg R. 2014. Tn5 transposase and tagmentation procedures for massively scaled sequencing projects. *Genome Res.* 24:2033–2040.

Purcell S, Neale B, Todd-Brown K, Thomas L, Ferreira MAR, Bender D, Maller J, Sklar P, De Bakker PIW, Daly MJ, et al. 2007. PLINK: A tool set for whole-genome association and population-based linkage analyses. *Am. J. Hum. Genet.* 81:559–575.

Purugganan MD. 2022. What is domestication? *Trends Ecol. Evol.* 37:663–671.

Qi Y, Xu J, Tian X, Bai Y, Gu X. 2017. The complete mitochondrial genome of *Hermetia illucens* (Diptera: Stratiomyidae). *Mitochondrial DNA Part B Resour.* 2:189–190.

Rambaut A, Drummond AJ, Xie D, Baele G, Suchard MA. 2018. Posterior summarization in Bayesian phylogenetics using Tracer 1.7. *Syst. Biol.* 67:901–904.

Rauw WM, Kanis E, Noordhuizen-Stassen EN, Grommers FJ. 1998. Undesirable side effects of selection for high production efficiency in farm animals: A review. *Livest. Prod. Sci.* 56:15–33.

Revell LJ. 2012. phytools: An R package for phylogenetic comparative biology (and other things). *Methods Ecol. Evol.* 3:217–223.

Rhode C, Badenhorst R, Hull KL, Greenwood MP, Bester-van der Merwe AE, Andere AA, Picard CJ, Richards C. 2020. Genetic and phenotypic consequences of early domestication in black soldier flies (*Hermetia illucens*). *Anim. Genet.* 51:752–762.

Rodrigues DP, Calado R, Pinho M, Rosário Domingues M, Antonio Vázquez J, Ameixa OMCC. 2022. Bioconversion and performance of Black Soldier Fly (*Hermetia illucens*) in the recovery of nutrients from expired fish feeds. *Waste Manag.* 141:183–193.

RStudio Team. 2019. RStudio: Integrated Development for R.

Saitou N, Nei M. 1987. The neighbor-joining method: a new method for reconstructing phylogenetic trees. *Mol. Biol. Evol.* 4:406–425.

Sheppard DC, Tomberlin JK, Joyce JA, Kiser BC, Sumner SM. 2002. Rearing methods for the black soldier fly (diptera: Stratiomyidae). *J. Med. Entomol.* 39:695–698.

Siddall ME, Trontelj P, Utevsky SY, Nkamany M, Macdonald KS. 2007. Diverse molecular data demonstrate that commercially available medicinal leeches are not *Hirudo medicinalis*. *Proc. R. Soc. B Biol. Sci.* 274:1481–1487.

Spielman D, Brook BW, Briscoe DA, Frankham R. 2004. Does inbreeding and loss of genetic diversity decrease disease resistance? *Conserv. Genet.* 5:439–448.

Spranghers T, Ottoboni M, Klootwijk C, Ovyn A, Deboosere S, De Meulenaer B, Michiels J, Eeckhout M, De Clercq P, De Smet S. 2017. Nutritional composition of black soldier fly (*Hermetia illucens*) prepupae reared on different organic waste substrates. *J. Sci. Food Agric.* 97:2594–2600.

Ståhls G, Meier R, Sandrock C, Hauser M, Šašić Zorić L, Laiho E, Aracil A, Doderović J, Badenhorst R, Unadirekkul P, et al. 2020. The puzzling mitochondrial phylogeography of the black soldier fly (*Hermetia illucens*), the commercially most important insect protein species. *BMC Evol. Biol.* 20:1–10.

Stamatakis A. 2014. RAxML version 8: A tool for phylogenetic analysis and post-analysis of large phylogenies. *Bioinformatics* 30:1312–1313.

Sun W, Yu H, Shen Y, Banno Y, Xiang Z, Zhang Z. 2012. Phylogeny and evolutionary history of the silkworm. 55:483–496.

Surendra KC, Tomberlin JK, Huis A Van, Cammack JA, Heckmann LL, Kumar S. 2020. Rethinking organic wastes bioconversion : Evaluating the potential of the black soldier fly (*Hermetia illucens* (L.)) (Diptera: Stratiomyidae) (BSF). *Waste Manag.* 117:58–80.

Tamura K, Subramanian S, Kumar S. 2004. Temporal Patterns of Fruit Fly (Drosophila) Evolution Revealed by Mutation Clocks. *Mol. Biol. Evol.* 21:36–44.

Thurmond J, Goodman JL, Strelets VB, Attrill H, Gramates LS, Marygold SJ, Matthews BB, Millburn G, Antonazzo G, Trovisco V, et al. 2019. FlyBase 2.0: The next generation. *Nucleic Acids Res.* 47:D759–D765.

Tomberlin JK, van Huis A. 2020. Black soldier fly from pest to “crown jewel” of the insects as feed industry: An historical perspective. *J. Insects as Food Feed* 6:1–4.

Verspoor RL, Haddrill PR. 2011. Genetic diversity, population structure and Wolbachia infection status in a worldwide sample of Drosophila melanogaster and D. simulans populations. *PLoS One* 6.

Wang MS, Thakur M, Peng MS, Jiang Y, Frantz LAF, Li M, Zhang JJ, Wang S, Peters J,

Otecko NO, et al. 2020. 863 Genomes Reveal the Origin and Domestication of Chicken. *Cell Res.* 30:693–701.

Xiang H, Liu X, Li M, Zhu Y, Wang L, Cui Y, Liu L, Fang G, Qian H, Xu A, et al. 2018. The evolutionary road from wild moth to domestic silkworm. *Nat. Ecol. Evol.* 2.

Zeder MA. 2015. Core questions in domestication research. *Proc. Natl. Acad. Sci. U. S. A.* 112:3191–3198.

Zhan Q, Zhao Y, Zhang S, Peng X. 2020. The complete mitochondrial genome of the soldier fly *Ptecticus aurifer*. *Mitochondrial DNA Part B Resour.* 5:660–661.

Zhan S, Fang G, Cai M, Kou Z, Xu J, Cao Y, Bai L, Zhang Y, Jiang Y. 2019. Genomic landscape and genetic manipulation of the black soldier fly *Hermetia illucens*, a natural waste recycler. *Cell Res.* 0:1–11.

Zhang C, Dong SS, Xu JY, He WM, Yang TL. 2019. PopLDdecay: A fast and effective tool for linkage disequilibrium decay analysis based on variant call format files. *Bioinformatics* 35:1786–1788.

Zhang C, Rabiee M, Sayyari E, Mirarab S. 2018. ASTRAL-III: Polynomial time species tree reconstruction from partially resolved gene trees. *BMC Bioinformatics*. 19:15–30.

Zhang J, Huang L, He J, Tomberlin JK, Li J, Lei C, Sun M, Liu Z, Yu Z. 2010. An artificial light source influences mating and oviposition of black soldier flies, *Hermetia illucens*. *J. Insect Sci.* 10:1–7.

Zhou F, Tomberlin JK, Zheng L, Yu Z, Zhang J. 2013. Developmental and waste reduction plasticity of three black soldier fly strains (Diptera: Stratiomyidae) raised on different livestock manures. *J. Med. Entomol.* 50:1224–1230.