ANIMAL GENETICS Immunogenetics, Molecular Genetics and Functional Genomics

doi: 10.1111/age.12961

Genetic and phenotypic consequences of early domestication in black soldier flies (*Hermetia illucens*)

C. Rhode*, R. Badenhorst*,†, K. L. Hull*, M. P. Greenwood*, A. E. Bester-van der Merwe*, A. A. Andere‡, C. J. Picard‡ and C. Richards†

*Department of Genetics, Stellenbosch University, Private Bag X1, Matieland 7602, South Africa. †AgriProtein Holdings Ltd, 1 Farnham Road, Guildford, SurreyGU2 4RG, UK. ‡Department of Biology, Purdue School of Science, Indiana University – University of Purdue Indianapolis, SL 306, 723 W Michigan Street, Indianapolis, IN 46202, USA.

Summary

The black soldier fly, *Hermetia illucens*, is an emerging biotechnological agent with its larvae being effective converters of organic waste into usable bio-products including protein and lipids. To date, most operations use unimproved commercial populations produced by mass rearing, without cognisance of specific breeding strategies. The genetic and phenotypic consequences of these commercial practices remain unknown and could have a significant impact on long-term population viability and productivity. The aim of this study was thus to assess the genetic and phenotypic changes during the early phases of colony establishment and domestication in the black soldier fly. An experimental colony was established from wild founder flies and a new microsatellite marker panel was developed to assess population genetic parameters along with the phenotypic characteristics of each generational cohort under captive breeding. The experimental colony was characterised by a small effective population size, subsequent loss of genetic diversity and rapid genetic and phenotypic differentiation between the generational cohorts. Ultimately, the population collapsed by the fifth generation, most likely owing to the adverse effect of inbreeding depression following the fixation of deleterious alleles. Species with *r*-selected life history characteristics (e.g. short life-span, high fecundity and low larval survival) are known to pose particular challenges for genetic management. The current study suggests that sufficient genetic and phenotypic variations exist in the wild population and that domestication and strain development could be achieved with careful population augmentation and selection during the early stages of colony establishment.

Keywords entomofarming, founder effects, genetic and phenotypic differentiation, inbreeding depression, selective breeding, selective sweep, vermiculture

Introduction

The black soldier fly, *Hermetia illucens*, is a cosmopolitan species that has spread across the world from its native neotropical habitat (Üstüner *et al.* 2003; Roháček & Hora 2013) and is considered neither a pest nor a vector of disease. It has been an organism of particular interest in recent years owing to the larvae's prolific role as decomposers of organic matter (Sheppard *et al.* 1994). As such, the black soldier fly's potential as a bioremedial agent, converting organic waste into usable animal protein (for conventional livestock and human consumption), lipids (for feeds/foods, biofuels and other industrial processes) and agricultural fertilisers has been studied (Sheppard *et al.* 1994; Diener *et al.* 2009; Banks *et al.* 2014; Lalande *et al.* 2014; Tschirner & Simon 2015). Consequently, this has led to the establishment of various black soldier fly production enterprises around the world, ranging from small-scale 'subsistence' or hobbyist operations to large industrial factories (Diener *et al.* 2009; Kenis *et al.* 2018).

With increasing concern for environmental sustainability, and global food and nutrition security, particularly in light of the growing human population, the availability of arable land for agriculture, diminishing fisheries stocks and unprecedented climate change, a shift to alternative food production practices is necessary (Nakagaki & Defoliart 1991; Rumpold & Schlüter 2013; Tomberlin *et al.* 2015). Vermiculture, and more generally entomofarming, has been proposed as one such alternative, with insects having high feed conversion ratios, low input costs and reduced space

Address for correspondence

C. Rhode, Department of Genetics, Stellenbosch University, Private Bag X1, Matieland 7602, South Africa.
E-mail: clintr@sun.ac.za

Accepted for publication 11 May 2020

© 2020 Stichting International Foundation for Animal Genetics, 51, 752–762

recruitments for production in comparison with traditional livestock (van Huis 2016). The black soldier fly is a prime candidate for insect domestication and production, owing to the availability of optimised mass rearing protocols (Sheppard *et al.* 2002) and its emerging associated industries. Therefore, it is envisioned that the number and scale of production facilities will increase in the near future.

Common practice for establishing a colony for mass rearing systems entails sourcing founder animals from the wild. This constitutes a population bottleneck, because it rarely encapsulates all available genetic variation and reduces the effective population size owing to the limited number of breeders. In the subsequent generations, when the reproductive cycle is closed, this leads to a further loss in genetic diversity due to the pronounced action of genetic drift and unintentional inbreeding. Simultaneously, selection for phenotypic traits associated with survival in the wild is relaxed, but selection for adaptation to the new artificial production environment might create a selective sweep, decreasing effective population size further. These commercial mass-reared populations might therefore be trapped in a cycle of genetic erosion. With these ongoing genetic processes there are normally associated unpredictable phenotypic changes (morphological, physiological and behavioural) that are also observed. At worst this might entail alterations to fitness traits that impact population viability, robustness and production capacity, due to the random fixation of deleterious alleles or inbreeding depression (Miyatake & Yamagishi 1999; Gilchrist *et al.* 2012; Baeshen *et al.* 2014). Genetic homogenisation of the population further hinders the evolutionary potential of the commercial colony, in terms of both response to artificial selection for enhanced production traits (including the development of specialised strains) and population resilience to unfavourable environmental permutations (Bartlett 1993; Baek *et al.* 2014). These effects are exacerbated by *r*-selected life history traits, commonly found amongst insects, including fast growth rates and short lifespans, high fecundity, but highly variable breeder contributions to offspring cohorts and mass larval mortality (Pianka 1970; Bufford & Daehler 2011). These characteristics have led to failures and challenges in the successful domestication, implementation of selective breeding programmes, and optimised production in some aquaculture species that exhibit similar life histories (e.g. Sass *et al.* 2010; Ozbay *et al.* 2014).

Therefore, to ensure the long-term sustainability of a commercial black soldier fly colony it is important to understand the micro-evolutionary processes that impact genetic diversity and phenotypic development during the early phases of domestication. Such knowledge, which is currently limited for black soldier flies, could be used to inform decision-making for colony management and directions for selective breeding and genetic improvement, and to ensure the genetic health and resilience of the colony. The aim of this study was therefore to evaluate genetic and phenotypic changes in the early phase of domestication (first four generations after establishment from a wild progenitor population). A microsatellite marker panel was developed to monitor genetic diversity and population phenotypic measurements were taken for the following traits, important in general fitness and production: clutch size, hatchability, pupae weight, percentage pupation, percentage eclosion, sex ratio and post-mating female longevity (percentage oviposition).

Materials and methods

Colony establishment, maintenance, sampling and DNA extraction

The experimental colony was established with wild prepupae (~2000 individuals) collected from Durban, South Africa (29.8587°S, 31.0218°E) and maintained for five discrete, successive generations (F0–F4; the species is semelparous). The standard husbandry protocol as described by Tomberlin *et al.* (2002) was followed. For each of the generations (F0–F4), 30 specimens were selected at random for genomic DNA extraction using the standard cetyltrimethyl ammonium bromide protocol described by Saghai-Maroof *et al.* (1984).

Molecular marker development

Hermetia illucens whole genome shotgun reads were retrieved from GenBank's Sequence Read Archive [accession nos: SRX265066 (male specimen) and SRX265065 (female specimen)]. The Sequence Read Archive files were converted to FastQ format using the SRA TOOLKIT (version 2.3.4-3) and *de novo* assembly was performed using CLC GENOMIC WORKBENCH (version 7.0.4). A k-mer size of 24 nucleotides for the assembly was selected and bubble size was set to 100 bp. All contigs shorter than 500 bp were removed from assembly. The remaining contigs were then used to search for tetra-nucleotide repeat motif microsatellites in MSATCOMMANDER (Faircloth 2008), specifically screening for a minimum of four repeats with at least two Cs or Gs, to avoid ATTT-like regions that are often troublesome in amplification and with sufficient flanking sequences for primer design. Specific primers for each fragment were designed in PRIMER3 (Untergasser *et al.* 2012) using default parameters.

KAPA2G Fast Multiplex PCR Kit (KAPA Biosystems) was used for PCR amplification in final volumes of 10 μl, with 0.8 μM of each primer, 0.2 mM of each dNTP and 50 ng DNA. Thermo-cycling parameters were as follows: initial denaturation at 95°C for 5 min, 35 cycles of 95°C for 15 s, T_A (annealing temperature) for 30 s, 72°C for 50 s and a final extension at 72°C for 90 s. Initially, gradient PCRs were performed to determine optimal annealing

temperatures for each primer pair. Thereafter, primers with consistent amplification and similar annealing temperatures were grouped for multiplex reactions. The forward primers were labelled with the ABI DS-33 standard dye-set (6-FAM™, VIC®, NED™, PET®) to facilitate multiplexing (Table S1). Capillary electrophoresis was done on ABI3730xl Genetic Analyser™ (Life Technologies) with GeneScan™ 500 LIZ® (Life Technologies) as internal size standard. Polymorphic detection and allele size scoring was done using GENEMAPPER® version 4.1 (Life Technologies).

Genetic analysis

Preliminary data quality assessments

Percentage polymorphism was noted, and genotypic data checked for the presence of genotyping errors owing to null alleles, allele stuttering and large allele drop-out in MICRO-CHECKER version 2.2.3 (Van Oosterhout *et al*. 2004) running 1000 permutations at the 5% nominal level. Null allele frequencies were estimated using the Brookfield 2 method (Brookfield 1996). Polymorphic information content and genotypic and allele frequencies were estimated with MICROSATELLITE TOOLKIT version 3.1.1 (Park 2001). LD between all pairs of loci was assessed in GENEPOP version 4.0 (10 000 dememorisations, 500 batches and 5000 iterations per batch; Rousset 2008) and neutrality was tested using an F_{ST}-outlier test in ARLEQUIN version 3.5.2 (20 000 simulations, 100 demes; Excoffier & Lischer 2010).

Genetic diversity statistics

For each generational cohort conformation to HWE was tested by Fisher's exact probability test in GENEPOP version 4.0 (10 000 dememorisations, 500 batches and 5000 iterations per batch). The number of alleles (A_N), observed heterozygosity (H_O) and unbiased expected heterozygosity (uH_E) were calculated in the MICROSATELLITE TOOLKIT version 3.1.1. Additional genetic diversity statistics were calculated in GENALEX version 6.502 (Peakall & Smouse 2006): the effective number of alleles (A_E) and Shannon's diversity information index (I). To avoid sampling bias, rarefied allelic richness (A_R) was calculated for each locus and each generation in HP-Rare (Kalinowski 2005). A Kruskal–Wallis test was performed to evaluate the significance of differences in the genetic diversity estimates between the generational cohorts.

Effective population size, census population size and estimation of population bottlenecks

Two methods were used to estimate effective population (N_E) size; the LD method (LD-N_E, with minimum allele frequency of 0.02) and the temporal method (temporal-N_E, using F0 as the progenitor for F1; F1 for F2; F2 for F3; and F3 for F4) under a random mating model in NEESTIMATOR version 2.0 (Do *et al*. 2014). Census population size (N_C) was obtained by directly counting all dead flies from the cage floor after each generation. The N_E/N_C ratio was calculated for each generational cohort using the LD-N_E point estimate. The Wilcoxon signed-rank test for excess heterozygosity (as evidence of a recent population bottleneck) was employed using the infinite alleles model, the stepwise mutation model and the two-phase model in BOTTLENECK version 1.2.02 (Piry *et al*. 1999). Analysis was performed using 1000 replications at the 5% nominal level and the two-phase model was composed of 70% stepwise mutation model and 30% infinite alleles model with a variance of 30.

Relatedness and inbreeding

Mean within-population pairwise relatedness (r) was computed at the 95% confidence interval (9999 permutations) for each generation according to the method described by Queller & Goodnight (1989) in GENALEX (Peakall & Smouse 2006). Inbreeding estimates described by locus-specific F_{IS} were estimated by the Weir & Cockerham (1984) method in GENEPOP.

Population genetic differentiation

To evaluate population differentiation, pairwise F_{ST} was calculated between all cohorts in GENALEX, with significant testing by permutation (999 replicates at the 1% nominal level) (Peakall & Smouse 2006). Additionally, pairwise Jost's D values (Jost 2008) were calculated between generations using the R package *DEMEtics* (Gerlach *et al*. 2010), with correction for sample size bias (reported as D_{est}) performed according to the method of Nei & Chesser (1983) and statistical significance determined by permutation (1000 bootstraps at the 1% nominal level). A Mantel test (Mantel 1967) was performed in GENALEX version 5.602 using pairwise F_{ST} and D_{est} values in correlation with generational 'distance' between the cohorts, with statistical significance determined through 999 permutations at the 5% nominal level.

Genetic clustering patterns amongst generations were evaluated using the multivariate discriminant analysis of principal components (DAPC), available in the R package *adegenet* (Jombart 2008). Prior to running the DAPC, cross-validation was performed to determine the optimal number of principal components to retain that allowed for the most accurate assignment of individuals (>80%) to specific genetic clusters.

Additionally, a hierarchical analysis of molecular variance (AMOVA) was performed in ARLEQUIN version 3.5.2, with 10 000 permutations to determine statistical significance. The hypothesis of genetic homogeneity among groups was evaluated, using an inter-generational grouping hypothesis

© 2020 Stichting International Foundation for Animal Genetics, **51**, 752–762

[generations separated into progenitor (F0) and captive bred (F1–F4)].

Phenotypic measurement and analysis

Phenotypic trait observations were recorded for each generational cohort as per Table S2, and analysed using various R packages. The *vegan* package was used for Non-Metric Dimensional scaling (NMDS) analysis (Dixon 2003) and boxplot and NMDS graphics were drawn using the *ggplot2* package (Wickham 2011). Wilcoxon Sum Rank test *P*-values displayed between generations on boxplots were calculated and visualised utilising the stat_compare_means () function in the *ggpubr* extension package for *ggplot2*. Spearman tests for detection of correlation between genetic estimates and phenotypic measurements across generations were conducted using the *Hmisc* package (Harrell 2019). Pairwise-P_{ST} values between generational cohorts were calculated using the *Pstat* package, using 1000 bootstraps to estimate 95% confidence intervals (Da Silva & Da Silva 2018). The pairwise P_{ST} estimates are analogous to pairwise F_{ST} and provide a means of comparison between the magnitude of phenotypic and genetic differentiation between groups (Baillie *et al.* 2016). A Mantel test (Mantel 1967) was performed in GENALEX version 5.602 using pairwise P_{ST} values in correlation with generational 'distance' between the cohorts, with statistical significance determined through 999 permutations at the 5% nominal level. Furthermore, to assess congruence between genetic and phenotypic differentiation, the correlations between pairwise P_{ST} and both pairwise F_{ST} and D_{est} values were computed using Spearman's rank test.

Results and discussion

Marker development

A total of 498 tetra-nucleotide loci were identified from next-generation sequence data available on NCBI. Of these 285 had sufficient flanking sequences for primer design, with primers having similar annealing temperatures and 40 loci were randomly selected for optimisation. Twenty-nine primer pairs were successfully optimised in PCR reactions and 10 (34.5%) were confirmed to be polymorphic with consistent amplification across the 150 specimens genotyped amongst the five generational cohorts. The average number of alleles ranged from two to 21 per marker, and genotypic patterns associated with null alleles, as well as high fixation indices, were detected at several loci. Nonetheless, the polymorphic information content ranged from 0.07 (*Hill25*, generation F4) to 0.90 (*Hill30*, generation F0; Table S3) and no significant signatures of selection or LD between loci were found. As such, this marker set was considered informative for standard population genetic inferences.

Population size, genetic diversity and differentiation across generations

There were significant differences in allelic richness (A_R), Shannon's index (I) and unbiased expected heterozygosity (uH_E) between generational cohorts (Kruskal–Wallis: $P < 0.05$); with a clear trend of reduced genetic diversity from the wild (F0) to the latter generations in captivity (Fig. 1). This is an expected result of an isolated population and has been observed in other captured insects under mass rearing (e.g. Miyatake 1998; Baek *et al.* 2014). Inbreeding coefficients (F_{IS}) varied temporally, all of which indicated elevated levels of homozygosity in all of the generations investigated. Unsurprisingly, deviations from HWE were observed across all generations (Table S3).

An increasing level of inbreeding is further supported by the relatedness estimates, showing incremental increases in consanguineous mating across the generational cohorts, with only the F0 having a mean negative population-wide relatedness coefficient (Fig. 2).

The significantly negative relatedness coefficients for the F0 (95% CI does not include 0), suggest that the original wild founder cohort was unintentionally sourced from differentiated populations; this is also supported by the low observed heterozygosity in relation to the expected heterozygosity for this generational cohort (Fig. 1), i.e. the Wahlund effect (Wahlund 1928). Although collected from the same geographical locality, temporary extinctions and recolonisation of flying insects in localised areas are not uncommon and this can lead to the mixture of individuals from different genetic backgrounds (e.g. Vandewoestijne *et al.* 1999; Okeyo *et al.* 2017). Black soldier fly populations tend to exhibit higher population structure on the local scale than on the regional scale (Park *et al.* 2017), suggesting mostly limited movement within local areas, but long-range migration does occur, and this could be facilitated by attractants from desirable substrates during resource-scarce periods (Nyakeri *et al.* 2017). However, there are few studies on black soldier fly migration patterns and genetic population structure in the wild, and none for South Africa currently.

The effects of this admixture can further be observed in the F1 generation that exhibited higher heterozygosity estimates (indicative of isolate breaking) and the increasing effective population size up to the F2 generation in both the LD- and temporal-N_E estimates (Table 1), when a genetic bottleneck was expected owing to the founder event (a bottleneck is however seen in the census population size, with a five times reduction in N_C from F0 to F1; Table 1). In fact, there is stronger evidence for a population expansion, a lingering artefact of isolate breaking (Table S4). Nonetheless, towards generation F3 and F4, the consequences of closing the reproductive cycle, creating an isolated population, are clearly observed with a sharp decrease in the LD-N_E to below the critical 50 value for a population at risk of

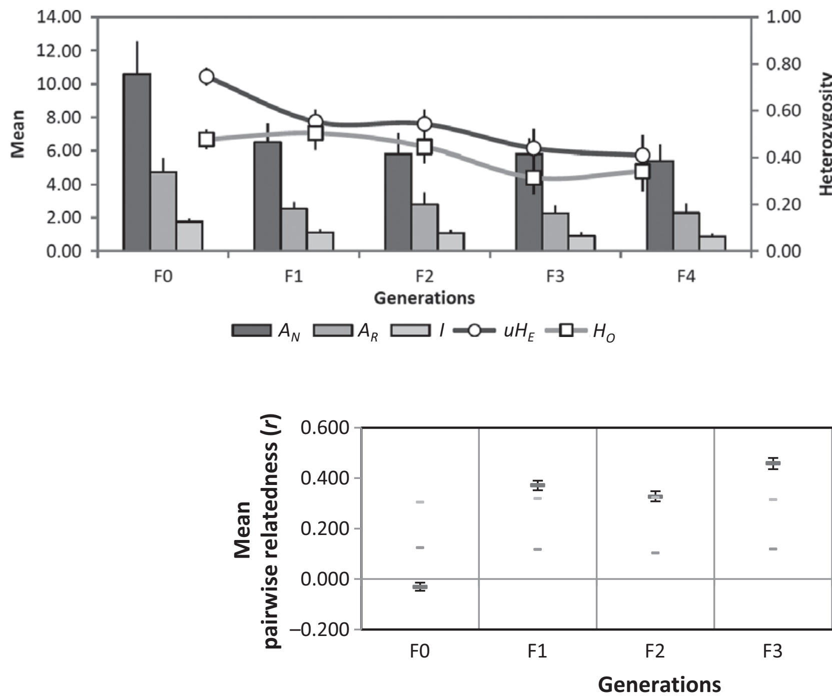

Figure 1 Summary of mean genetic diversity estimates for each sampling population using 10 microsatellite loci. A_N, Number of alleles; A_R, allelic richness (rarefied); I, Shannon's Index; uH_E, unbiased expected heterozygosity; H_O, observed heterozygosity (a detailed account is given in Table S3).

Figure 2 Mean within-population pairwise relatedness estimates computed as per Queller & Goodnight (1989). Mean relatedness is represented in black bars, with 95% confidence intervals about the mean given by the error bars. The 'free floating' grey bars indicate upper and lower confidence intervals (95%) testing the hypothesis for significant difference between mean relatedness across the generational cohorts.

extinction (Frankham *et al.* 2014), accompanied by a census population size decrease in generation F4. This led to a population collapse in the F5 with all of the pupae failing to eclose, most likely the result of accelerated inbreeding depression caused by synergistic epistasis of randomly fixed deleterious alleles across multiple loci (Domínguez-García *et al.* 2019).

Luikart *et al.* (2010) argue the LD-N_E estimate is more powerful than temporal-N_E for the early detection of bottlenecks, and as such the N_E/N_C ratios are given in terms of the LD-N_E point estimate. Stark, unpredictable generational fluctuations in the N_E/N_C ratio for animals with *r*-selected life history characteristics (high fecundity, variance in reproductive and offspring survival) are common, but highlight the risk of such populations rapidly losing genetic diversity, compromising population viability (Luikart *et al.* 2010; Ferchaud *et al.* 2016). This is evident in the current study with N_E/N_C ratios ranging from 0.162 (F1) to 0.004 (F3) and the systematic loss of diversity across the generations (Table 1), further contributing to the population crash in the F5 generation.

A further consequence of the small N_E is the rapid differentiation of the generational cohorts from the original wild F0 owing to the pronounced effects of genetic drift. Pairwise F_{ST} and D_{est} values between all generations were statistically significant (Table S5), and significant correlation between genetic distance and generational separation (Mantel test: $F_{ST} - R = 0.721$, $P = 0.026$; $D_{est} - R = 0.268$, $P = 0.025$) was found. This was further supported by the

Table 1 Estimates of effective population size (N_E) calculated using the LD and temporal methods.

Method	F0	F1	F2	F3	F4
LD	46.2 (34.1–67.8)	78.4 (34.1 to ∞)	∞ (95.4 to ∞)	7.7 (5.2–10.9)	9.2 (6.1–13.7)
Temporal	4.0 (2.6–5.9)	11.6 (6.0–23.0)	9.4 (5.0–17.5)	28.7 (11.5–126.9)	
Census population size	2434	484	3517	1938	689
N_E/N_C	0.019	0.162	NA	0.004	0.013

95% confidence intervals are indicated in brackets. The census population for each cohort (N_C) and the N_E/N_C ratio are also reported.

hierarchical AMOVA (Table 2), with significant differentiation found between the generational groups (wild progenitor vs. captive bred) and between generations within groups (explaining ~5 and 8% of genetic variance respectively). Multivariate genetic clustering (DAPC, Fig. 3) separated the groups into the wild progenitor population (F0) and the subsequent captive bred generations (F1–F4) along the *x*-axis. Along the *y*-axis, generations F1 and F2 clustered separately from generations F3 and F4.

Temporal instability of population structure has been observed specifically in *r*-selected species where stochastic factors, often owing to small N_E, have a significant impact on allelic variances across generations (e.g. Chapman *et al.* 2002; Hoffman *et al.* 2004; Okeyo *et al.* 2017; Rhode *et al.* 2017). Furthermore, as currently reported for the black soldier fly, gradual increases in the level of differentiation over time were similarly observed for *Drosophila subobscura* when a wild population was used to establish a laboratory colony (Santos *et al.* 2012). Nowak *et al.* (2007) also reported that under an artificial selection regime, the harlequin fly (*Chironomus riparius*), showed a rapid divergence within the first few generations of captive breeding. Although the present study did not apply a specific directed selection pressure, the move from a natural habitat to an enclosed artificial habitat can exert an selective pressure and selection for adaptation to the novel environment could have a significant impact on population differentiation in black soldier flies (Mignon-Grasteau *et al.* 2005).

Phenotypic trends and correlations with genetic parameters

Insects that live in highly variable and seasonal environments often exhibit phenotypic plasticity in life history trait regulation, which buffers populations against abrupt environmental permutations (van Dyck & Wiklund 2002). The black soldier fly might have an inherited capacity for phenotypic plasticity granted its colonisation ability, spreading across the globe from its native range in South America (Nyamukondiwa *et al.* 2010). Furthermore, plasticity in black soldier flies has been observed in the utilisation of different feed substrates (Tomberlin *et al.* 2002; Zhou *et al.* 2013). This could explain the large phenotypic ranges and shifts in life-cycle progression traits across generations (percentage eclosion, percentage oviposition, percentage pupation and percentage hatchability; Fig. 4a, Table S6).

However, the population under investigation is also undergoing specific genetic processes associated with the founder effect. The low hatchability in the F1, and the sharp drop in the percentage pupation and percentage eclosion from F0 to F1 is perhaps not surprising and likely to be the result of a selective sweep enabling rapid adaptation to the new artificial environment (Pélissié *et al.* 2018). The percentage oviposition is resultantly higher for the F1 generation as surviving females capitalise on their adaptive fitness, possibly further bolstered by hybrid vigour owing to the admixture effect as discussed earlier. Hatchability, percentage eclosion and percentage pupation recovered in the F2, most likely owing to the gained adaptive fitness. Interestingly, percentage pupation and percentage eclosion further increased or remain fairly stable in the F3 and F4, whilst hatchability and percentage oviposition declined. The drastic reduction in N_E and the increasing genetic homogenisation of the population in the F3 and F4 generations hints at different mechanisms of inbreeding depression for these life history traits. Inbreeding depression in rates of pupation and eclosion is likely under the control of a dominance mechanism, and when the population was purged from recessive deleterious alleles (during generation F1), the optimal phenotypic values were rapidly restored in the following generations. This is supported by the significant, negative correlation between percentage pupation and observed heterozygosity (Table S7), where an increased genome-wide heterozygosity probably relates to the

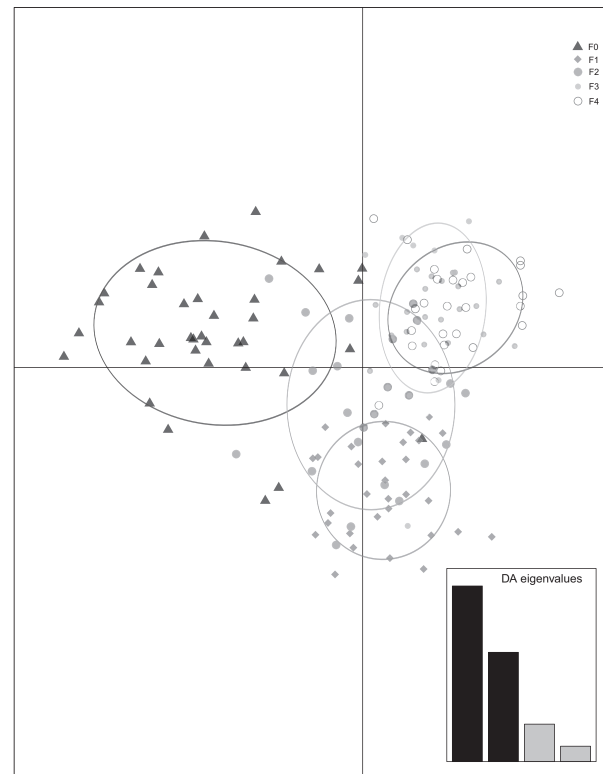

Figure 3 Discriminant analysis of principal components plot of generational cohorts of *Hermetia illucens*, with each generation represented by different shapes and shades of grey.

Table 2 Hierarchical AMOVA for the evaluation of the inter-generational grouping hypothesis of the *Hermetia illucens* system.

Hypothesis tested	Source of variation	Variation (%)	Fixation index
Inter-	generational	Amongst groups	4.94
	$F_{ST} = 0.13$*		
	Amongst generations within groups	7.92	$F_{SC} = 0.08$*
Within		$F_{CT} = 0.05$**	
groups	87.14		

*Statistical significance at α = 0.01.
**Statistical significance at α = 0.05.

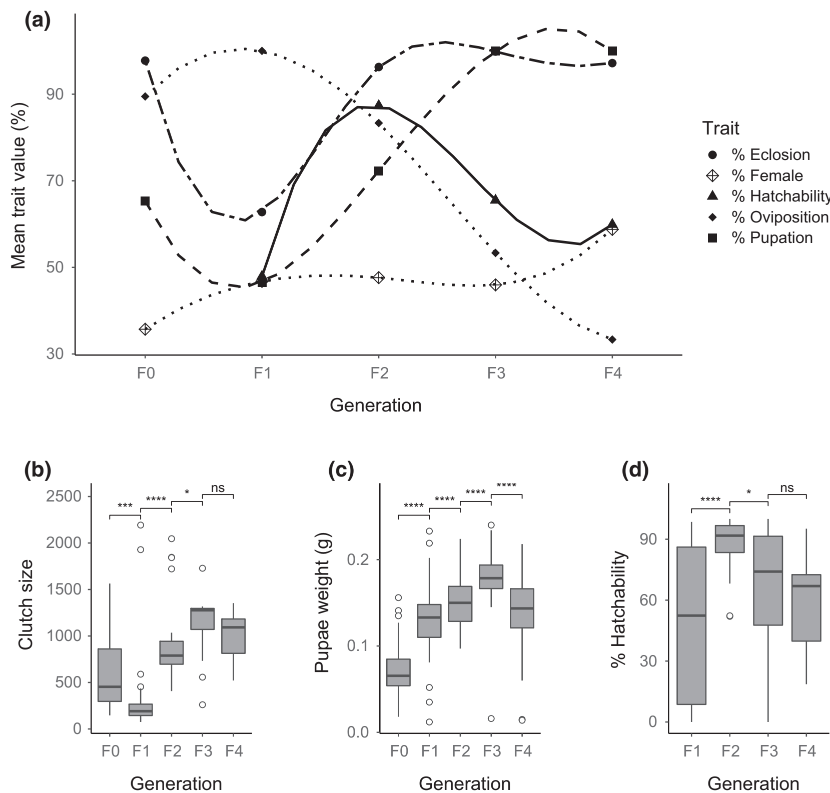

Figure 4 (a) Trends in the measurements of life-cycle progression traits (percentage eclosion, percentage oviposition, percentage pupation, percentage hatchability) and sex bias (percentage female). Boxplots display distributions for generational measures of egg clutch size (b), pupae weight (c) and percentage hatchability (d). White circles represent outliers in the data. Crossbars denote statistical comparison of group means by application of a Wilcoxon Sum Rank test; n.s., non-significant; *significant at $\alpha = 0.05$; ***significant at $\alpha = 0.001$; ****significant at $\alpha = 0.0001$.

presence of more deleterious alleles at loci responsible for regulating these traits. On the contrary, inbreeding depression in oviposition and hatchability might be under the influence of an overdominance mechanism and is particularly supported by the significantly positive correlation between observed heterozygosity and percentage oviposition (Table S6; Charlesworth & Willis 2009).

The percentage oviposition and clutch size were significantly negatively correlated (Table S7). At first this might seem unexpected and counter-intuitive; however, it is known that clutch size is a plastic trait where some insects might regulate clutch sizes based on feed and substrate availability (e.g. Aluja *et al.* 2019). Thus, for the current experiment, with constant fixed resources, the fewer females that oviposited the more available resources there were *per capita*, leading to females that might have up-regulated their clutch sizes. Alternatively, noting that clutch sizes follow a similar phenotypic trend across the generations (Fig. 4b) to percentage pupation and percentage elcosion, the disparate mechanisms of inbreeding depression might also explain the observation, as discussed above. The prevalence of female adults is noted to have gradually increased over time; however, it was within the range of previously reported sex ratios (Tomberlin *et al.* 2002).

The quantitative measures of clutch size, pupae weight and hatchability (Fig. 4b–d) showed statistically significant temporal changes within the colony (Wilcoxon Sum Rank test, $P < 0.05$), with the largest shifts in mean values noted over the first three generations of isolated rearing (F0–F3; Table S6). This was supported by significant phenotypic differentiation across all generations based on P_{ST} estimates (Tables S8–S10), with the greatest magnitude of sequential generational differentiation generally detected between F0 and F1 [clutch size, 0.713 (0.049–0.951); pupae weight, 0.984 (0.975–0.991); percentage hatchability, 0.923 (0.839–0.966)], supporting the hypothesis of a 'hard' selective sweep. Whereas these estimates mirror pairwise genetic F_{ST} and D_{est} values between many generational comparisons, pairwise P_{ST} values do not reliably detect the same extent of differentiation between later generational cohorts for clutch size and percentage hatchability. Furthermore, the fairly high phenotypic differentiation vs. neutral genetic differentiation might suggest that these traits are under directional selection (Baillie *et al.* 2016; da Silva & da Silva. 2018). Shimoji & Miyatake (2002) observed similar phenotypic fluctuations in laboratory-reared populations (from wild progenitors) of sweet potato weevil (*Euscepes postfasciatus*), but only reported this magnitude of fluctuations at the later phases of isolated breeding. These authors had a starting census population that was three times greater than that reported for the black soldier fly in this study. The rapid fluctuation of the phenotypic values might thus be a function of small N_E leading to stochastic fluctuations in allele frequencies underlying these complex traits. Mantel tests for P_{ST} correlations with generational distance were non-

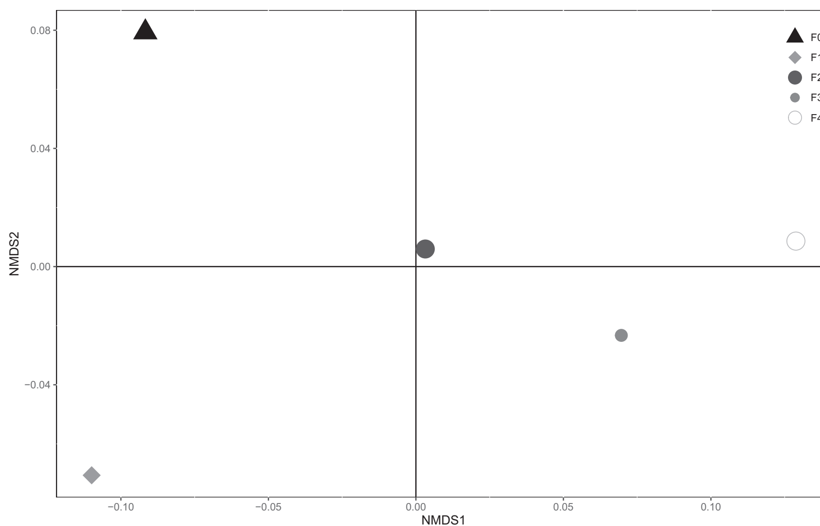

Figure 5 Non-metric dimensional scaling analysis of discrete generations of a *H. illucens* colony based on mean generational phenotypic measures (clutch size, pupae weight, % adult females, percentage eclosion, percentage pupation and percentage oviposition) and genetic diversity estimates (F_{IS}, allelic richness and mean pairwise relatedness). Each generation is represented by a unique shape and shade of grey.

significant, and there was also no significant correlation between phenotypic differentiation and genetic differentiation with either F_{ST} or D_{est} (Spearman's correlations, $P > 0.05$), supporting the non-directionality of phenotypic differentiation. It can also be noted that multiallelic loci, such as microsatellites, might often deflate estimates of genetic differentiation (Balloux & Lugon-Moulin 2002), whereas most loci underlying traits are more likely to be biallelic, representing single nucleotide variations. Lastly, the role of plasticity in phenotypic differentiation cannot be excluded and this may inflate the P_{ST} estimates.

Nonetheless, the NMDS analysis incorporating both phenotypic and genetic parameter estimates yielded a successive distribution of generational cohorts along the primary explanatory axis (Fig. 5). This stepwise pattern of separation between the generations recapitulates and supports relationships retrieved by DAPC analysis (Fig. 3), where these temporal cohorts are similarly drawn out along the primary explanatory axis. A similar pattern of rapid and parallel phenotypic and genotypic change has been reported for the dronefly (*Eristalis tenax*) under captive rearing (Francuski *et al.* 2014).

Conclusions

The black soldier fly is an insect that shows great promise for waste bioremediation and as an alternative source of protein (and other bio-products) for human and animal utilisation. As with conventional livestock and aquaculture, animal production systems (including entomofarming) are most sustainable, efficient and profitable with effective genetic management and breeding strategies. Species with *r*-selected life history traits, however, pose particular challenges, as observed in the current study. Founder effects, including selective sweeps, further confounded by high variance in reproductive success lead to small N_E across generations after the closure of the reproductive cycle in captivity. As a result, there is excessive loss of genetic diversity, which has implications for population viability and evolutionary potential for response to artificial selection for the development of specialist production strains. These genetic effects have consequences for phenotypic development in both fitness traits, such as fertility and larval survival, and production traits like larval growth rates. The current study suggests that there is sufficient genetic and phenotypic variation in black soldier populations in South Africa to develop domesticated strains with enhanced phenotypic performance for production traits. A future population genetic assessment of the wild populations could, however, add value to decision-making on sourcing animals for founder commercial populations. From the current results it is clear that the early phases of domestication must be managed carefully with a combination of population augmentation of the commercial colonies coupled with careful selective breeding for favourable phenotypes to ensure long-term colony genetic health and productivity. Ideally, the founder populations should be established from a broad genetic background to maximise genetic diversity that should serve as a buffer against potential adverse founder effects, as has been observed in generations F0–F2. An N_E of at least 500 is normally considered a genetically stable population. This can be maintained by regularly augmenting the commercial colony through the introduction of wild flies at predetermined generational intervals. During the early phases of colony establishment this might be vital for counteracting inbreeding depression and preventing colony collapse. However, augmentation with wild flies could be counter-productive in latter generations in terms of genetic gains made through selective breeding. Future investigations may therefore explore potential cross-breeding or rotational mating strategies to mitigate this and exploit hybrid vigour for relevant production traits.

Acknowledgements

The authors would like to thank AgriProtein Holding Ltd for the use of the facilities and providing partial funding, and the undergraduate researchers at IUPUI who contributed to the early screening of microsatellite markers. The National Research Foundation South African is acknowledged for funding this research (reference no. CSRP170506229933).

Data availability

All data, not confined to the main text, are available in the Supplementary Information.

References

Aluja M., Birke A., Díaz-Fleischer F. & Rull J. (2019) Phenotypic plasticity in clutch size regulation among populations of a potential invasive fruit fly from environments that vary in host heterogeneity and isolation. *Bulletin of Entomological Research* **109**, 169–77.

Baek M.J., Yoon T.J., Kang H.J. & Bae Y.J. (2014) Biological and genetic characteristics of *Glyptotendipes tokunagai* (Diptera: Chironomidae) on the basis of successive rearing of forty-two generations over seven years under laboratory conditions. *Environmental Entomology* **43**, 1406–18.

Baeshen R., Ekechukwu N.E., Toure M., Paton D., Coulibaly M., Traoré S.F. & Tripet F. (2014) Differential effects of inbreeding and selection on male reproductive phenotype associated with the colonization and laboratory maintenance of Anopheles gambiae. *Malaria Journal* **13**, 19.

Baillie S.M., Muir A.M., Hansen M.J., Krueger C.C. & Bentzen P. (2016) Genetic and phenotypic variation along an ecological gradient in lake trout *Salvelinus namaycush*. *BMC Evolutionary Biology* **16**, 219.

Balloux F. & Lugon-Moulin N. (2002) The estimation of population differentiation with microsatellite markers. *Molecular Ecology* **11**, 155–65.

Banks I.J., Gibson W.T. & Cameron M.M. (2014) Growth rates of black soldier fly larvae fed on fresh human faeces and their implication for improving sanitation. *Tropical Medicene and International Health* **19**, 14–22.

Bartlett A.C. (1993) Maintaining genetic diversity in laboratory colonies of parasites and predators. In: *Applications of Genetics to Arthropods of Biological Control Significance* (Ed. by S.K. Narang, A.C. Bartlett & R.M. Faust), pp 133–45. CRC Press, Boca Raton, FL.

Brookfield J.F.Y. (1996) A simple new method for estimating null allele frequency from heterozygote deficiency. *Molecular Ecology* **5**, 453–5.

Bufford J.L. & Daehler C.C. (2011) Life history strategies. In: *Encyclopedia of Biological Invasions* (Eds. by D. Simberloff & M. Rejmanek) pp. 437–41). University of California Press, Los Angeles, CA.

Chapman R.W., Ball A.O. & Mash L.R. (2002) Spatial homogeneity and temporal heterogeneity of red drum (*Sciaenops ocellatus*) microsatellites: effective population sizes and management implications. *Marine Biotechnology* **4**, 589–603.

Charlesworth D. & Willis J.H. (2009) The genetics of inbreeding depression. *Nature Reviews Genetics* **10**, 783.

Da Silva S.B. & Da Silva A. (2018) Pstat: an R package to assess population differentiation in phenotypic traits. *R Journal* **10**, 447–54.

Diener S., Zurbrügg C. & Tockner K. (2009) Conversion of organic material by black soldier fly larvae: establishing optimal feeding rates. *Waste Management Research* **27**, 603–10.

Dixon P. (2003) Computer program review VEGAN, a package of R functions for community ecology. *Journal of Vegetation Science* **14**, 927–30.

Do C., Waples R., Peel D., Macbeth G., Tillett B.J. & Ovenden J.R. (2014) NeESTIMATOR v2: re-implementation of software for the estimation of contemporary effective population size (Ne) from genetic data. *Molecular Ecology Research* **14**, 209–14.

Domínguez-García S., García C., Quesada H. & Caballero A. (2019) Accelerated inbreeding depression suggests synergistic epistasis for deleterious mutations in *Drosophila melanogaster*. *Heredity* **123**, 709–22.

Excoffier L. & Lischer H.E.L. (2010) Arlequin suite ver 3.5: a new series of programs to perform population genetics analyses under Linux and Windows. *Molecular Ecology Research* **10**, 564–7.

Faircloth B.C. (2008) MSATCOMMANDER: detection of microsatellite repeat arrays and automated, locus-specific primer design. *Molecular Ecology Research* **8**, 92–4.

Ferchaud A.L., Perrier C., April J., Hernandez C., Dionne M. & Bernatchez L. (2016) Making sense of the relationships between Ne, Nb and Nc towards defining conservation thresholds in Atlantic salmon (*Salmo salar*). *Heredity* **117**, 268.

Francuski L., Djurakic M., Ludoški J., Hurtado P., Pérez-Bañón C., Ståhls G. & Milankov V. (2014) Shift in phenotypic variation coupled with rapid loss of genetic diversity in captive populations of *Eristalis tenax* (Diptera: Syrphidae): consequences for rearing and potential commercial use. *Journal of Economic Entomology* **107**, 821–32.

Frankham R., Bradshaw C.J. & Brook B.W. (2014) Genetics in conservation management: revised recommendations for the 50/500 rules, Red List criteria and population viability analyses. *Biological Conservation* **170**, 56–63.

Gerlach G., Jueterbock A., Kraemer P., Deppermann J., & Harmand P. (2010) Calculations of population differentiation based on G_{ST} and D: forget G_{ST} but not all of statistics!. *Molecular Ecology* **19**, 3845–52.

Gilchrist A.S., Cameron E.C., Sved J.A. & Meats A.W. (2012) Genetic consequences of domestication and mass rearing of pest fruit fly *Bactrocera tryoni* (Diptera: Tephritidae). *Journal of Economic Entomology* **105**, 1051–6.

Harrell F.E. Jr (2019) Package 'Hmisc'. *CRAN2018*, **2019**, 235–6.

Hoffman E.A., Schueler F.W. & Blouin M.S. (2004) Effective population sizes and temporal stability of genetic structure in *Rana pipiens*, the northern leopard frog. *Evolution* **58**, 2536–45.

van Huis A. (2016) Edible insects are the future? *Proceedings of the Nutrition Society* **75**, 294–305.

Jombart T. (2008) Adegenet: a R package for the multivariate analysis of genetic markers. *Bioinformatics* **24**, 1403–5.

Jost L. (2008) G_{ST} and its relatives do not measure differentiation. *Molecular Ecolology* **17**, 4015–26.

Kalinowski S.T. (2005) HP-RARE 1.0: a computer program for performing rarefaction on measures of allelic richness. *Molecular Ecology Notes* **5**, 187–9.

Kenis M., Bouwassi B., Boafo H. *et al.* (2018) Small-scale fly larvae production for animal feed. In: *Edible Insects in Sustainable Food Systems* (Ed. by A. Halloran, R. Flore, P. Vantomme & N. Roos), pp 239–61. Springer, Cham.

Lalander C.H., Fidjeland J., Diener S., Eriksson S. & Vinnerås B. (2014) High waste-to-biomass conversion and efficient *Salmonella* spp. reduction using black soldier fly for waste recycling. *Agronomy for Sustainable Development* **35**, 261–71.

Luikart G., Ryman N., Tallmon D.A., Schwartz M.K. & Allendorf F.W. (2010) Estimation of census and effective population sizes: the increasing usefulness of DNA-based approaches. *Conservation Genetics* **11**, 355–73.

Mantel N. (1967) The detection of disease clustering and a generalized regression approach. *Cancer Research* **27**, 1183–5.

Mignon-Grasteau S., Boissy A., Bouix J. *et al.* (2005) Genetics of adaptation and domestication in livestock. *Livestock Production Science* **93**, 3–14.

Miyatake T. (1998) Genetic changes of life history and behavioural traits during mass-rearing in the melon fly, *Bactrocera cucurbitae* (Diptera: Tephritidae). *Researches on Population Ecology* **40**, 301–10.

Miyatake T. & Yamagishi M. (1999) Rapid evolution of larval development time during mass-rearing in the melon fly, *Bactrocera cucurbitae*. *Research in Population Ecology* **41**, 291–7.

Nakagaki B.J. & Defoliart G.R. (1991) Comparison of diets for mass-rearing *Acheta domesticus* (Orthoptera: Gryllidae) as a novelty food, and comparison of food conversion efficiency with values reported for livestock. *Journal of Economic Entomology* **84**, 891–6.

Nei M. & Chesser R. (1983) Estimation of fixation indices and gene diversities. *Annuls of Human Genetics* **47**, 253–9.

Nowak C., Vogt C., Diogo J.B. & Schwenk K. (2007) Genetic impoverishment in laboratory cultures of the test organism *Chironomus riparius*. *Environmental Toxicology and Chemistry* **26**, 1018–22.

Nyakeri E.M., Ogola H.J., Amimo F.A. & Ayieko M.A. (2017) Comparison of the performance of different baiting attractants in the egg laying activity of the black soldier fly (*Hermetia illucens* L.). *Journal of Entomology and Zoology Studies* **5**, 1583–6.

Nyamukondiwa C., Kleynhans E. & Terblanche J.S. (2010) Phenotypic plasticity of thermal tolerance contributes to the invasion potential of Mediterranean fruit flies (*Ceratitis capitata*). *Ecological Entomology* **35**, 565–75.

Okeyo W.A., Saarman N.P., Mengual M. *et al.* (2017) Temporal genetic differentiation in *Glossina pallidipes* tsetse fly populations in Kenya. *Parasites & Vectors* **10**, 471.

Ozbay G., Reckenbeil B., Marenghi F. & Erbland P. (2014) *Eastern Oyster (*Crassostrea virginica*) Aquaculture and Diversity of Associated Species*, pp. 1–54. Dover, DE: Delaware State University, Department of Agriculture and Natural Resources.

Park S. (2001) The Excel microsatellite toolkit. http://animalgenomics.ucd.ie/sdepark/ms-toolkit.

Park S., Choi H., Choi J.Y. & Jeong G. (2017) Population structure of the exotic black soldier fly, *Hermetia illucens* (Diptera: Stratiomyidae) in Korea. *Korean Journal of Environment and Ecology* **31**, 520–8.

Peakall R. & Smouse P.E. (2006) GENALEX 6: genetic analysis in Excel. Population genetic software for teaching and research. *Molecular Ecology Notes* **6**, 288–95.

Pélissié B., Crossley M.S., Cohen Z.P. & Schoville S.D. (2018) Rapid evolution in insect pests: the importance of space and time in population genomics studies. *Current Opinion in Insect Science* **26**, 8–16.

Pianka E.R. (1970) On *r*- and *K*-selection. *The American Naturalist* **104**, 592–7.

Piry S., Luikart G. & Cornuet J.M. (1999) BOTTLENECK: a computer program for detecting recent reduction in the effective population size using allele frequency data. *Journal of Heredity* **90**, 502–3.

Queller D.C. & Goodnight K.F. (1989) Estimating relatedness using genetic markers. *Evolution* **43**, 258–75.

Rhode C., Bester-van der Merwe A.E. & Roodt-Wilding R. (2017) An assessment of spatio-temporal genetic variation in the South African abalone (*Haliotis midae*), using SNPs: implications for conservation management. *Conservation Genetics* **18**, 17–31.

Roháček J. & Hora M. (2013) A northernmost European record of the alien black soldier fly *Hermetia illucens* (Linnaeus, 1758) (Diptera: Stratiomyidae). *Cas Slezskeho Zemskeho Muz* **62**, 101–6.

Rousset F. (2008) GenePop007: a complete re-implementation of the genepop software for Windows and Linux. *Molecular Ecology Resources* **8**, 103–6.

Rumpold B.A. & Schlüter O.K. (2013) Potential and challenges of insects as an innovative source for food and feed production. *Innovative Food Science & Emerging Technologies* **17**, 1–11.

Saghai-Maroof M.A., Soliman K.M., Jorgensen R.A. & Allard R.W. (1984) Ribosomal DNA spacer-length polymorphisms in barley: Mendelian inheritance, chromosomal location, and population dynamics. *Proceedings of National Academy of Sciences* **81**, 8014–8.

Santos J., Pascual M., Simões P., Fragata I., Lima M., Kellen B., Santos M., Marques A., Rose M.R. & Matos M. (2012) From nature to the laboratory: the impact of founder effects on adaptation. *Journal of Evolutionary Biology* **25**, 2607–22.

Sass G.G., Cook T.R., Irons K.S., McClelland M.A., Michaels N.N., O'Hara T.M. & Stroub M.R. (2010) A mark-recapture population estimate for invasive silver carp (*Hypophthalmichthys molitrix*) in the La Grange Reach, Illinois River. *Biological Invasions* **12**, 433–6.

Sheppard D.C., Newton G.L., Thompson S.A. & Savage S. (1994) A value added manure management system using the black soldier fly. *Bioresource Technology* **50**, 275–9.

Sheppard D.C., Tomberlin J.K., Joyce J.A., Kiser B.C., & Sumner S.M. (2002) Rearing methods for the black soldier fly (Diptera: Stratiomyidae). *Journal of Medical Entomology* **39**, 695–8.

Shimoji Y. & Miyatake T. (2002) Adaptation to artificial rearing during successive generations in the West Indian sweetpotato weevil, *Euscepes postfasciatus* (Coleoptera: Curculionidae). *Annals of the Entomological Society of America* **95**, 735–9.

Tomberlin J.K., Sheppard D.C. & Joyce J.A. (2002) Selected life-history traits of black soldier flies (Diptera: Stratiomyidae) reared on three artificial diets. *Annals of the Entomological Society of America* **95**, 379–86.

Tomberlin J.K., van Huis A. & Benbow M.E. *et al.* (2015) Protecting the environment through insect farming as a means to produce protein for use as livestock, poultry, and aquaculture feed. *Journal of Insects as Food and Feed* **1**, 307–9.

Tschirner M. & Simon A. (2015) Influence of different growing substrates and processing on the nutrient composition of black soldier fly larvae destined for animal feed. *Journal of Insects as Food and Feed* **1**, 249–59.

Untergasser A., Cutcutache I., Koressaar T., Ye J., Faircloth B.C., Remm M., & Rozen S.G. (2012) Primer3 – new capabilities and interfaces. *Nucleic Acids Research* **40**, 1–12.

Üstüner T., Hasbenli A. & Rozkošný R. (2003) The first record of *Hermetia illucens* (Linnaeus, 1758) (Diptera, Stratiomyidae) from the Near East. *Stud Dipterologica* **10**, 181–5.

Van Dyck H. & Wiklund C. (2002) Seasonal butterfly design: morphological plasticity among three developmental pathways relative to sex, flight and thermoregulation. *Journal of Evolutionary Biology* **15**, 216–25.

Van Oosterhout C., Hutchinson W.F., Wills D.P.M. & Shipley P. (2004) MicroChecker: software for identifying and correcting genotyping errors in microsatellite data. *Molecular Ecology Notes* **4**, 535–8.

Vandewoestijne S., Nève G. & Baguette M. (1999) Spatial and temporal population genetic structure of the butterfly *Aglais urticae* L. (Lepidoptera, Nymphalidae). *Molecular Ecology* **8**, 1539–43.

Wahlund S. (1928) Zuzammensetzung von populationen und korrelation-serscheiunungen von standpunkt der vererbungslehre aus betrachtet. *Hereditas* **11**, 65–106.

Weir B.S. & Cockerham C.C. (1984) Estimating F-Statistics for the analysis of population structure. *Evolution* **38**, 1358–70.

Wickham H. (2011) ggplot2. *WIRE. Computational Statistics* **3**, 180–5.

Zhou F., Tomberlin J.K., 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. *Journal of Medical Entomology* **50**, 1224–30.

Supporting information

Additional supporting information may be found online in the Supporting Information section at the end of the article.

Table S1 Characterisation of molecular markers utilised in the assessment of experimental generations of *Hermetia illucens*

Table S2 Descriptive methodologies for the estimation and measurement of all phenotypic traits assessed within this study

Table S3 Genetic diversity indices per microsatellite marker for *H. illucens* across experimental generations and one commercial generation: polymorphic information content (PIC); average number of alleles (A_N); allelic richness (A_R); Shannon's information index (I); observed heterozygosity (H_O); expected heterozygosity (H_E); fixation index/inbreeding coefficient (F_{IS}) [*indicates deviation from HWE ($P < 0.01$)]; and null allele frequencies (Fr (Null)), as well as the standard error (SE) for each mean estimate

Table S4 Wilcoxon's signed rank test significance values (P) for heterozygous deficiency and excess for all cohorts, assuming three mutation models, namely the infinite alleles model, the two-phase model and the stepwise mutation model

Table S5 Genetic differentiation estimates for generational cohorts of *H. illucens*

Table S6 Mean generational measures of all phenotypic traits included in this study

Table S7 Spearman correlation rank testing applied to generational means of genotypic diversity estimates (PIC, A_N, A_R, I, H_o, r) and phenotypic measures (percentage hatchability, percentage eclosion, percentage oviposition, percentage pupation, percentage females, clutch size and pupae weight)

Table S8 Pairwise P_{ST} values for the clutch size trait for all generations assessed, with 95% confidence interval bounds in parentheses

Table S9 Pairwise P_{ST} values for the pupae weight trait for all generations assessed, with 95% confidence interval bounds in parentheses

Table S10 Pairwise P_{ST} values for the percentage hatchability trait for all generations assessed, with 95% confidence interval bounds in parentheses

© 2020 Stichting International Foundation for Animal Genetics, **51**, 752–762