Abstract

A major ongoing research effort seeks to understand the behavior, ecology and control of the spotted lanternfly (SLF) (*Lycorma delicatula*), a highly-invasive pest in the U.S. and South Korea. These insects undergo four nymphal stages (instars) before reaching adulthood, and appear to shift host plant preferences, feeding, dispersal and survival patterns, anti-predator behaviors, and response to traps and chemical controls, with each stage. However, categorizing SLF lifestage is challenging for the first three instars, which have the same coloration and shape, because no comprehensive allometric datasets exist. We present a dataset of body mass and length for SLF nymphs throughout a growing season and compare our results with published ranges of instar body lengths based on small samples. An ontogenetic allometric analysis found that SLF nymph body mass scales isometrically with body length (exponent c = 3.05 [2.95,3.15]). An analysis using two clustering methods also revealed that first through third instar body mass and length fell into distinct clusters (Dyar’s rule), supporting using these two
metrics to stage nymphs during a single growing season. The ranges for 2nd and 3rd instars were not consistent between our results and those from earlier studies for diverse locations. Using previously-published data, we also found that tarsal claw and arolia (adhesive footpad) dimensions scale in proportion to body length and mass, respectively, indicating that adhesive ability does not decrease with age, as posited in some previous studies. Conversely, mouthpart dimensions do not correlate with body length, consistent with predictions that these features should reflect preferred host plant characteristics rather than body size. We suggest extending these methods to study how SLF instar development depends on factors such as hatch date, host plant, temperature, and geographic location, using citizen scientist networks to collect morphometric data for a wide range of locations and environmental conditions.

Introduction

The spotted lanternfly (SLF) (*Lycorma delicatula*) is a planthopper native to south Asia that has become a highly invasive pest in the U.S. and South Korea. SLFs feed intensively on phloem from a wide variety of plants and trees, stressing the hosts as well as promoting the growth of sooty mold (1). Because SLFs threaten significant economic damage to agricultural crops, native trees, and landscape plants, a large ongoing research effort seeks to understand their development, physiology, behavior and ecology to inform methods for mitigation and control (2–4). In this study, we discuss how clustering methods can be applied to measurements of the body mass and size of immature SLFs (nymphs) in order to improve the determination of SLF lifestage and to study the allometric scaling of these SLF footpart and mouthpart dimensions with body morphometrics. We begin by explaining how these issues are relevant to a wide variety of topics in SLF research.

After emerging, SLFs develop through five life stages separated by molting: four wingless nympha instars and the much larger and winged adult stage. The first three instars have
identical black and white coloration and morphology and differ only in size, while the 4th instars are readily identified by their distinctive red, black and white coloring. Many studies of SLF behavior, ecology, and phenology have relied on determination of the nymphal stage (instar determination) in order to track how life stage influences ecology and choice of host plants (1,2), dispersal patterns (3–6), locomotor behaviors such as climbing and jumping (7,8), phenology and activity (9), spectral preferences (10), attraction to chemicals (11), and effectiveness of various trapping methods (12). Other studies have relied on estimated life stages in exploring how foot, mouth part and antenna morphology change during development (7,13,14) to provide information on how these factors influence feeding, adhesion and locomotion throughout the insect’s life cycle. Thus, instar determination methods for determining the life stage of a given specimen collected in the field are useful and important in many contexts.

Fig 1. Photograph of 1st, 2nd, 3rd, and 4th instar spotted lanternfly nymphs, with the double-headed arrow showing the definition of body length, L. (scale bar = 10 mm).

In spite of this growing interest, only a few previous studies have reported actual data on body dimensions for use in determining instar size, and none have reported body mass. (Table 1, Fig 1) The earliest study reported only mean body lengths for each stage in China (15). Park and colleagues (16) measured body lengths for 1st through 4th instars in South Korea, although it was not stated whether specimens used for measurements were raised in the laboratory with
known life stage or collected from the wild and the instar stage estimated from size. Jang et al. (10) reported only body lengths for just 2nd instars captured in the field in South Korea. Dara et al. (17) reported the ranges of body lengths measured for 1st through 4th instar nymphs collected in Pennsylvania. None of these previous studies provided statistical data to guide the classification of new datasets for instar determination. Furthermore, studies of other insect species have shown that the size ranges for each instar can depend on factors such as date of emergence, diet (18), host plants, temperatures, and environment (e.g., laboratory vs field-raised)(19). Indeed, prior research has indicated that SLF nymphs develop and survive differently when reared with different diets in the field (19,20), at different temperatures (21), and artificial conditions (i.e., enclosures or laboratory conditions) (1,16), but these studies did not consider how these factors affected instar morphometrics.

Table 1. Body length (mm) of spotted lanternfly nymphs from this study and earlier work.

<table>
<thead>
<tr>
<th>Life stage</th>
<th>Body length (mm)</th>
<th>Ref. (17) [range]</th>
<th>Ref. (10) mean ± SD</th>
<th>Ref. (16) mean ± SE</th>
<th>Ref. (15) mean</th>
</tr>
</thead>
<tbody>
<tr>
<td>1st instar</td>
<td>4.24 ± 0.24, N = 54</td>
<td>[3.6, 4.4], N = 12</td>
<td>Not measured</td>
<td>3.9 ± 0.2, N = 43</td>
<td>4</td>
</tr>
<tr>
<td>2nd instar</td>
<td>6.67 ± 0.48, N = 30</td>
<td>[5.1, 6.4], N = 10</td>
<td>6.0 ± 0.5, N = 17</td>
<td>5.7 ± 0.7, N = 62</td>
<td>7</td>
</tr>
<tr>
<td>3rd instar</td>
<td>9.30 ± 0.72, N = 72</td>
<td>[6.9, 9.4], N = 12</td>
<td>Not measured</td>
<td>8.9 ± 0.4, N = 23</td>
<td>10</td>
</tr>
</tbody>
</table>
In other insect and arthropod species, instar determination has often relied on Dyar’s rule: the finding that nymphal body sizes tend to be distributed among distinct clusters because each instar stage attains a maximum body size before molting (22). The use of Dyar’s rule ideally involves directly measuring the frequency distribution of one or more morphometric measures for each instar using specimens with known molting status (e.g., from measuring molted head capsule dimensions) (23). However, instar determination should be possible without knowledge of molting status if the number of developmental stages is known in advance, the morphometric data is uniformly sampled across all life stages, and its frequency distribution is partitioned into distinct clusters (24). The last approach is especially useful for SLFs, which have proven challenging to raise in the laboratory so that life stage can be directly monitored (20,21), and which we have observed to have flaccid cast exoskeletons that do not provide useful sizing information after molting.

In this study, we report measurements of mass and body length for spotted lanternfly nymphs along with clustering results for these specimens. With these data, we were able to test whether spotted lanternfly nymph body mass scales allometrically with length, as has been found for insects and other arthropods (25), and how the dimensions of other body parts scale with overall body size. We also describe a citizen science project in which we propose to collect morphometric data for SLFs at all stages of development from a wide variety of locations and conditions with the goal of understanding the variation in SLF development and providing open access to this data and analysis methods for other studies.
Methods

Insect collection and morphometrics

Healthy, intact SLF nymphs were collected from *Ailanthus altissima* trees and wild grape vines (*Vitis vinifera*) in southeastern Pennsylvania (40°00'30.2"N 75°18'22.0"W) from May through August, 2021. We collected and measured specimens without prescreening for size to avoid sampling bias. Because SLF are identified as an invasive species in Pennsylvania, all specimens were euthanized by freezing (26). Morphometric data were measured post-mortem after thawing for 15 min to preserve tissue hydration and morphology using an analytical balance (Explorer, Ohaus, Parsippany, NJ US) to measure mass, M (accuracy ± 0.4 mg). Body length, L, was defined as snout-caudal length and measured using ImageJ (27) to ± 0.05 mm from digital micrograph images that included a scale bar in the same plane. (Fig 1, Table 1, Table 2) A total of N = 156 1st through 3rd instars and N = 68 4th instars were collected and used for clustering and fitting.

Table 2. Morphometric data for spotted lanternfly nymph mass and body length.

<table>
<thead>
<tr>
<th>Life stage</th>
<th>N</th>
<th>Mean mass ± SD (mg)</th>
<th>Mean length ± SD (mm)</th>
<th>CV length</th>
</tr>
</thead>
<tbody>
<tr>
<td>1st instar</td>
<td>54</td>
<td>3.0 ± 1.2</td>
<td>4.2 ± 0.2</td>
<td>5.8%</td>
</tr>
<tr>
<td>2nd instar</td>
<td>30</td>
<td>12.0 ± 5.5</td>
<td>6.7 ± 0.5</td>
<td>7.3%</td>
</tr>
<tr>
<td>3rd instar</td>
<td>61</td>
<td>33.4 ± 12.0</td>
<td>9.3 ± 0.7</td>
<td>7.8%</td>
</tr>
<tr>
<td>4th instar</td>
<td>68</td>
<td>59.2 ± 13.0</td>
<td>11.7 ± .8</td>
<td>6.4%</td>
</tr>
</tbody>
</table>

The 4th instars were identified by coloration, while 1st, 2nd and 3rd instars were classified using fits of the mass vs length to a 3 component Gaussian mixture model. (N = number specimens)
Comparison with other studies

We performed Google Scholar searches using the keywords spotted lanternfly and Lycorma delicatula, yielding over 600 references. The most relevant approximately 100 papers were used to perform repeated forward and reverse citation searches to find morphometric data for spotted lanternfly nymphs. This resulted in the identification of four papers with additional values of body length (10,15–17). One study reported footpart data for adhesive pad (aroilium) and tarsal claw dimensions and mouthpart dimensions for the labium and the stylet, which is used to pierce plant surfaces for feeding; we used this footpart morphometric data to compute the area of the arolium (14). (S1 Appendix)

Data analysis

Data analysis was performed using MATLAB version R2021a with the curve fitting and statistics and machine learning toolboxes (Mathworks, Natick MA USA); MATLAB functions are referred to using italicized names. All data and code required to reproduce all results and figures discussed here are accessible at https://doi.org/10.6084/m9.figshare.19287389.v1.

All 4th instars were identified by their red, black and white coloring. Length and mass data for all specimens with black and white coloration consistent with 1st through 3rd instars were standardized before clustering by converting them into z-scores (i.e., zero mean and standard deviation = 1). The standardized data were fit to a three component Gaussian Mixture Model using fitgmdist (covariance type = full, shared covariance = false), then sorted into three components (clusters) using cluster in MATLAB to reflect the known number of instar stages in the dataset. We also partitioned only the length data for the first, second and third instars into three clusters using the Gaussian Mixture Model and kmeans for k-means clustering.
To determine whether the body mass vs length data obey an allometric scaling law with exponent $c$, $M = A L^c$, we also fit the log-transformed data to $\log M = c \log L + A$ using linear regression (MATLAB \texttt{fitlm} command). Because there were only four or five distinct data points (one for each life stage studied) from morphometric data analyzed from previous studies, we only fitted their dependence on body length or mass using linear regression, rather than a generalized scaling law. Pearson's correlation coefficients for some datasets were also computed using \texttt{corrcoef}.

**Results**

**Instar determination**

The results of our morphometric measurements are shown in Fig 2 along with clustering data using the GMM model for mass vs body length; summary statistics are given in Tables 1 and 2. The data were sorted into identical clusters using GMM clustering for either mass and length or length-only and k-means for length-only.

**Allometry**

Our log-transformed data agreed well with the allometric scaling law, $\log M = c \log L + A$, with scaling exponent $c = 3.05 [2.95, 3.15]$ (R-squared = 0.95; F-statistic vs. constant model: $4.1 \times 10^3$, p-value $= 3 \times 10^{-145}$). (Fig 3) The fit residuals were symmetrically distributed about the mean apart from 4 outliers for which the residuals were $> 3$ standard deviations from the zero mean (2.5% of 156 total points). Fitting the data with these residual outliers omitted resulted in similar fit results ($< 1.5\%$ difference in the exponent and R-squared.)
Fig 2. Clustering of SLF nymph mass vs length data. Measurements are shown as points, shaded ellipses show the 95% CI for each cluster based on the Mahalanobis distance, and x’s indicate cluster centroids. (Note that the shaded ellipse for the first instar cluster is covered by datapoints.) The solid red line at top shows allometric scaling law fit to the data. Solid horizontal lines at bottom show the lengths reported for each instar reported in the previous studies (Table 1).
Fig 3. (A) Mass vs length data shown on a log-log scale with the allometric fit. (B) Fit residuals vs log L. (C) Histogram of fit residuals.
The distance between tarsal claw tips, TCT, for all life stages of SLFs (14) also scaled isometrically with body length: $TCT = (78.8 \pm 5.5) L$ (R-squared = 0.986, $P = 0.0007$) with zero intercept ($P = 0.12$). (Fig 4A) By contrast, the arolium area, $A_{adh}$, for nymphs from the same study varied in direct proportion to body mass: $A_{adh} = (2.80 \pm 0.11) \times 10^3 \text{ micron}^2/\text{mg} \ M$ (R-squared = 0.999, $P = 0.002$), with intercept not significantly different from zero ($P = 0.71$); we did not include the adult value in the fit because it was significantly smaller than the value for any nymphal stages. Mouthpart dimensions (labium and stylet length) did not exhibit a significant correlation with body length across the different instar stages (labium: Pearson’s $R = 0.91$, $p = 0.086$; stylet: Pearson’s $R = 0.74$, $P = 0.26$).
Fig 4. Comparison of spotted lanternfly A) tarsal claw tip width, B) arolium area, and C) mouthpart lengths from (14) with body mass (A) and length (B,C) from this study. (Fit lines from linear regression in A, B; all error bars = SEM)
Discussion

The results of this study lead to several conclusions. First, we found that the overall distribution of 1st to 3rd instar size data sizes matched our expectations from Dyar’s Rule of three distinct clusters, with no overlap between clusters in the approximate size ranges expected for these life stages from previous studies. While collecting body mass data facilitates allometric modeling, body lengths alone provided the same information as both mass and length as far as clustering was concerned. This indicates that easy-to-perform specimen body length measurements indeed should indeed be sufficient for instar determination. Practical applications of these findings include that indication that clustering methods with large datasets of spotted lanternfly body lengths should facilitate determining the likely developmental stage of individual specimens in future studies. The datasets and clustering code provided as supplemental materials enables other researchers to replicate and extend these results either to estimate the lifestage of new specimens based on our data, or to perform clustering on their own measurements.

Second, Fig 1 shows that there is considerable variation between different studies for both the reported means and ranges of instar body lengths. Our estimated ranges for 1st and 4th instar clusters overlapped with the ranges previously reported in the literature. On the contrary, the ranges for 2nd and 3rd instars were not consistent either between past studies or between our study and most published ranges. These differences could be due to a variety of factors, including different environmental conditions, differences in population morphology, or variations in study design (e.g., different sampling methods used to collect different life stages). The diversity of reported values for instar sizes suggests that it would be interesting to compile and analyze morphometric data as a function of lifestage for SLFs raised under a variety of
circumstances to understand how much of this variation is real and how much due merely to study design.

Only the body length and mass ranges for 3rd and 4th instars estimated in our study displayed considerable overlap. One possible explanation for this is that the nymphs become more sexually dimorphic as they develop. While female adults are larger than males (21), we have not found any data on size variations in nymphal stages. Due to the red coloration of 4th instar nymphs, however, this overlap in ranges does not cause a logistical challenge for properly categorizing these life stages.

Third, these data also provide insights into the ontogenic allometric scaling of spotted lanternfly nymph body metrics. The body mass vs length for all SLF nymphs obeys an allometric scaling law with an exponent consistent with the value of 3 expected for isometric scaling. This indicates that the geometric shape of SLF nymphs is approximately constant prior to the adult stage, and consequently that their biomechanics can be described by a single mechanical model, appropriately scaled to the size of the instar in question.

Fourth, these data can be used to create new syntheses of existing research for greater insight into the biology of these insects. For example, Kim et al. (7) hypothesized that earlier SLF instars should be more easily dislodged by wind than later nymphs due to their smaller arolia, an idea with implications for how dispersal and control should depend on lifestage. However, we found that arolium area (14) is directly proportional to mass for SLF nymphs (Fig 4A). This dependence agrees with the empirical relationship ($A_{adh} \propto M^{0.02}$) found for hemipterans, anurans and across many species with sizes spanning seven orders of magnitude in weight (28), and is consistent with the scaling relationship predicted for constant maximum adhesive stress between the arolium and surface. This implies that earlier SLF instars do not have
disproportionately small adhesive pads compared to later instars and that therefore one would not expect them to be more easily dislodged from host plants. Indeed, an analysis of the data for the number of first instars on trees per day as a function of wind speed from (7) (Fig 5) does not find a significant correlation between these measures (Pearson R = 0.43, P = 0.28), contrary to what would be expected if these early instar nymphs were weakly adhered. The results found here indicate the reported monotonic increase in SLF falling-combing cycle period with advancing date of the year (7) is likely due to factors other than arolium development, such as the isometric growth of tarsal claws and the findings in (14) of increased wrinkling of the arolia surface and a larger terminal sticky lip in adult SLFs relative to nymphs. These findings for adhesion are also relevant because of the crucial role transportation plays in the dispersal of SLFs, which are known to travel long distances by clinging to vehicles and shipping containers (29).

Fig 5. Plot of data from Fig. 3 (7) for number of first instars observed on trees per day vs ambient wind speed.

By contrast, the stylet and labium lengths do not correlate significantly with body length for SLF nymphs, even though both mouthparts are ≥3 times longer in adults than nymphs (14). This finding is consistent with the expectation that stylet length is correlated with preferred host plant
tissue characteristics (30), as opposed to insect size, given reports from the literature indicate that SLF nymphs only feed on herbaceous and non-woody parts of plants (e.g., shoots, stems and leaves) while adults are able to feed on bark-covered trunks (7,9,16,17,31).

To further explore all of these issues going forward, we plan to create a citizen science project to encourage the collection of additional data, thus addressing the limitation that gathering data for SLFs over a wide range of geographic locations and other conditions is challenging for any single research group. Past citizen scientist projects have proved significant, cost-effective contributions to the study of insects (32) including invasive species (33–35), for early detection, surveillance and monitoring, determining range expansion, and community-based control efforts. Because the distinctive appearance of SLF makes their detection and identification feasible for nonexperts, these efforts include multiple ongoing initiatives aimed at mapping SLF distributions (36,37) containing and eradicating SLF populations (38), and determining which animals prey on SLFs (39). We believe that there is a similar need for a platform that provides a standardized protocol for measuring the size of SLF nymphs and recording their environmental data, and then providing open access to an archive of the collected results. In combination with the code provided here, these datasets could be used to perform clustering and other analyses to estimate instar stage and correlate morphometric data with other factors. This effort could take advantage of the large workforce and geographic reach of citizen science to facilitate the study of correlations between instar morphometrics and factors such as date of first emergence, molt schedule, geographic location, host plant, and temperature. These data could play a role in defining fitness benchmarks needed for interpreting other field biology data.

Acknowledgements

We wish to thank Chengpei Li and Aaron Xu for assistance in collecting specimens.
References

14. Avanesyan A, Maugel TK, Lamp WO. External morphology and developmental changes of larsal tips and mouthparts of the invasive spotted lanternfly, Lycorma delicatula (Hemiptera:


37. Penn State Extension. Have you seen a Spotted Lanternfly? Let’s check! [Internet]. Penn State Extension. [cited 2022 Feb 10]. Available from: https://extension.psu.edu/have-you-seen-a-spotted-lanternfly
