Measuring Mass: Variation among 3,161 species of Canadian Coleoptera and the prospects of a mass

registry for all insects

- 6 Jingchan Hu^{1,2}, Mikko Pentinsaari¹, Paul D.N. Hebert^{1,2}
- ¹ Centre for Biodiversity Genomics, University of Guelph, Guelph, Ontario, Canada
- Department of Integrative Biology, University of Guelph, Guelph, Ontario, Canada
-

-
- Corresponding Author:
- 13 Mikko Pentinsaari¹
- Centre for Biodiversity Genomics, University of Guelph, 50 Stone Road East, Guelph, Ontario,
- N1G 2W1, Canada
- Email address: mpentins@uoguelph.ca
-

Abstract

- Although biomass values are critical for diverse ecological and evolutionary analyses, they are
- unavailable for most insect species. Museum specimens have the potential to address this gap,
- 22 but the variation introduced by sampling and preservation methods is uncertain. This study
- quantifies variation in the body mass of 3,161 species of Canadian Coleoptera, employing the
- Barcode Index Number system for their discrimination, a critical requirement for the inclusion of
- groups where the taxonomic impediment prevents the assignment of specimens to a Linnaean
- species. By validating the reproducibility of measurements and evaluating the error introduced by operational complexities such as curatorial practice and the loss of body parts, it demonstrates
- 28 that museum specimens can speed the assembly of a mass registry. The results $\frac{1}{\sqrt{2}}$ ther indicate
- 29 that congeneric species of Coleoptera generally have limited variation in mass, so a genus-level
- identification allows prediction of the body mass of species that have not been weighed or
- 31 measured. Building on the present results, the construction of a mass registry for all insects is
- 32 feasible.
-

Introduction

-
- Body mass is a key property of organisms which impacts factors ranging from metabolic rate to
- 37 community structure, foraging behavior, and predator-prey dynamics (Peters, 1986; Chown &
- Gaston, 2010; Smith et al., 2016). Comprehensive body mass registries are available for
- mammals (Jones et al., 2009), fishes (Froese & Pauly, 2021), and birds (Dunning, 2008), but

- similar information is lacking for insects despite their abundance and ecological importance.
- Three characteristics of insects have impeded its assembly: 1) high species diversity, 2) variable
- curatorial practices, and 3) small size requires access to precision balances (Braun et al., 2009;
- Chown & Gaston, 2010; Gilbert, 2011; Knapp, 2012). These barriers to direct mass
- measurements have led many studies to employ estimates derived from body length, even for
- groups with very divergent body plans (Rogers, Hinds & Buschbom, 1976). Despite its lack of
- precision (Gowing & Recher, 1984; Johnston & Cunjak, 1999), this approach has been widely
- applied due to its simplicity (Ulrich, 2007; Greve et al., 2018; Richard, Tallamy & Mitchell,
- 2019). Aside from the fact that direct measurements of body mass for arthropods are uncommon,
- 49 existing data are difficult to access because there is no structured data repository (Chown $\&$ Gaston, 2010).
-
- The construction of a mass registry for insect species would benefit studies that currently depend
- on imprecise surrogates, facilitating comparisons across groups with differing morphology.
- Aggregating data from all insect orders and other arthropods, which are typically studied
- independently, would advance understanding of mass variation and its evolutionary trajectories
- across lineages (Ulrich, 2007; Chown & Gaston, 2010). Museum specimens has been proposed
- as a resource to assemble mass data for insects without new sampling effort (Gilbert, 2011).
- However, to assess the quality of the mass data resulting from their analysis, the impact of
- varying curatorial and preservation methods requires investigation.
-
- With over 360,000 described species, Coleoptera is one of the most diverse orders of insects.
- 62 Occurring in both aquatic and terrestrial environments, it includes some of the largest and
- smallest insects with its component taxa spanning eight orders of magnitude in mass (Chown &
- Gaston, 2010). These factors make Coleoptera an ideal group for developing approaches to
- support the construction of a mass registry for all insects. The present study targets the
- Coleoptera of Canada, a fauna of nearly 9,000 species (Brunke et al., 2019), many possessing a
- DNA barcode record on BOLD (Ratnasingham & Hebert, 2007). The mass data gathered in this
- study provide a strong basis for comparison with previous surveys (Chislenko, 1981; Novotny &
- Kindlmann, 1996; Ulrich, 2007). As well, because these values derive from specimens with
- DNA barcodes, it begins to develop the information on mass variation needed to advance both
- metabarcoding and eDNA analyses.
-
- This study details variation in the body mass of 3,161 species of Coleoptera based on the
- 74 analysis of museum specimens. It evaluates the impacts of humidity, tissue loss, and curatorial
- 75 variables on mass. It also examines the extent of variation in mass among taxonomic lineages,
- work which indicates that phylogenetic constraints are strong enough for the current data to
- allow mass estimation for most Canadian Coleoptera. Finally, this study considers how best to
- expand from the current registry that includes records for a few thousand species to one with
- coverage for all insect species.

Materials & Methods

BINs as a species proxy

85 DNA barcoding employs sequence variation in a 658 bp segment of the cytochrome q oxidase subunit I gene (COI) as a basis for specimen identification and species discovery in animals

- (Hebert et al., 2003). The BIN system clusters these COI sequences into molecular operational
- 88 taxonomic units (Floyd et al., 2002) that correspond closely with Linnaean species. Each BIN is
- 89 assigned a unique alphanumeric that serves as a species proxy (Ratnasingham & Hebert, 2013).
- For example, about 90% of all species in the well-studied European Coleoptera fauna show
- perfect correspondence with BINs (Pentinsaari, Hebert & Mutanen 2014; Hendrich et al., 2015).
- Because they provide a taxonomic assignment for undescribed species (Brunke et al. 2019;
- Brunke et al. 2021), this study employed BINs to structure data collection. While substantial
- efforts were also made to assign each BIN to a Linnaean species, this was not always possible
- because of both the lack of taxonomic specialists for some families and difficulties in resolving
- synonymies and cryptic species. As a result, we employ the BIN count as the best estimator of
- the number of species examined in this study.
-

Body mass data

 Specimens were available for 3,161 BINs of Canadian Coleoptera. They represented 1,100 of the 2,008 genera and 96 of the 111 beetle families known from Canada (Bousquet et al., 2013). Most were morphologically identified to a genus (3,156 BINs) and many to a species (2,719 BINs to 2,389 recognized species). The specimens were obtained from sampling programs coordinated by the Centre for Biodiversity Genomics (CBG) at the University of Guelph and are stored in its voucher collection. Specimens missing major body segments (head, abdomen) were not weighed. 107 In total, 3,744 specimens were analyzed, meaning a single specimen was weighed for most \equiv S. 108 However, 2–5 specimens were weighed for 334 BINs, and, for these taxa, a mean mass was 109 calculated.

-
- Specimens fell into three main curatorial categories (ethanol, pointed, pinned). Specimens from
- 70% ethanol were first air-dried and then weighed repeatedly until the mass measurement
- stabilized. Specimens on points were unmounted using 70% ethanol, dried, and weighed. Pinned
- specimens were weighed on their pin and the pin mass was subtracted. When mass variation
- among pins of the size used on a specimen exceeded 12.5% of its overall weight (Gilbert, 2011),
- it was unpinned and weighed directly.
-
- 118 The mass of small specimens (~2,800 representing 2,500 BINs) was quantified to the nearest
- 0.0001 mg using a high-precision balance (Mettler Toledo™ XP6U), while the ~950 larger

- 120 specimens $(>10 \text{ mg})$ were weighed to the nearest 0.1 mg using a less sensitive instrument
- (Mettler Toledo™ MS104S). Up to three significant figures were recorded. The BIN, taxonomic
- assignment, and mass of each analyzed specimen are provided in a supplementary document
- (Data S1).
-

Data description and distribution analyses

All analyses were performed in R version 3.6.3 (R Core Team, 2020). Mass values were log-

- transformed before further analysis, and the interquartile range (IQR) was used as a measure of
- variance. A two-sided D'Agostino test was employed to evaluate skewness in the data, and an
- Anscombe-Glynn test to assess kurtosis. One-way ANOVAs using family, subfamily, and
- generic assignments as variables were used to assess mass variation at different levels in the
- taxonomic hierarchy using respective groups containing two or more quantified BINs. A nested
- ANOVA was also used to examine variation partitioning in the 50 families and 449 genera that were nested with two or more subgroups. Because taxonomic groups with low species diversity
- tend to show less variation in mass (Chown & Gaston, 2010), a separate analysis was performed
- to ascertain how variation in mass was partitioned in the most diverse groups of Canadian
- Coleoptera. In particular, a one-way ANOVA examined 65 genera with mass data for 10 or more
- BINs, while a nested ANOVA examined the six families with mass data for >100 BINs. To
- 139 quantify the extent of mass that could be partitioned at each taxonomic level, ω 2 values were
- calculated for all ANOVAs.
-

Pin variation and reproducibility assessment

-
- 144 A pinne is not easily separated from \ddot{H} , creating a complexity because the pin can
- 145 outweigh the specimen. Gilbert (2011) proposed a workaround that involves estimating pin mass
- from key parameters (material, shape, size) before subtracting this value from the total weight to
- produce a mass value for the insect. Because the CBG employs insect pins from a single supplier, this source of variation was readily assessed. The mean and standard deviation in both
- diameter and mass was determined for 100 pins of each size. Because there was no overlap in
- diameter among different pin sizes, the size associated with each specimen could be determined,
- allowing its mass to be subtracted.
-
- The consistency in determinations of pin size and of body mass was assessed by comparing mass values for 120 specimens weighed in 2014 and again in 2018. The congruence in net mass values was examined using a paired-sample t-test.
-
- To evaluate short-term variation in mass, 20 specimens were weighed daily for a week when
- variation in humidity was pronounced. As well, 50 specimens were examined to determine the
- reduction in mass caused by the loss of a leg.

Results

Mass variation in Canadian Coleoptera

Measurements for 3,744 beetles representing 3,161 BINs revealed their mass varied by more

166 than five orders of magnitude (0.002 $\left(\frac{1}{\sqrt{2}}\right)$ 7 mg) (Fig 1). Among BINs with a species assignment,

Ptiliola kunzei (BOLD:ACI8875) and *Ptiliolum fuscum* (BOLD:AAM7677) possessed the lowest

mass (0.0056 mg). However, six BINs in the same subfamily (Ptiliinae) weighed less, and a

specimen identified to the genus *Nanosella* (BOLD:ADH5266) had the lowest mass (0.0024

mg). The largest species was *Hydrophilus triangularis* (Hydrophilidae, BOLD:AAQ2470) at 797

171 mg. The median mass of all species was \sim 1.3 mg, represented in the data by species such as

Bembidion nitidum (Carabidae, BOLD:AAD2752) and *Dichelotarsus piniphilus* (Cantharidae,

 BOLD:AAH0933). Considering all BINs, the mass distribution approximated a lognormal 174 distribution with strong kurtosis $(z = -7.39, p = 1.49-13)$ but insignificant skew $(z = 1.96, p = 1.49-13)$

- 0.051).
-

Much of the variation in mass among species was linked to their higher taxonomic assignments

178 (family, subfamily, genus) (Table 1. a-c). In fact, ω^2 values indicated that 90% of the mass

variation could be explained by higher taxonomic placement with 55% of the variation at the

family level, 20% at the subfamily level, and 15% at the genus level. Because of these

relationships, variation in mass among congeneric species was typically limited (Fig 2). In all,

for the 519 genera where two or more BINs were examined, the variance, measured by IQR, had

183 a median of 0.163 log_{10} (mg), which is a 1.4-fold difference and translates to $+/- 20\%$ divergence

 from the median for the genus. Cases of extreme variation where the larger members of the 185 genus were as twice as massive as the median $(IQR > 0.6)$ were only observed in nine genera.

Reliability of specimen mass measurements

Diameter measurements allowed the discrimination of each pin size as differences among pins of

190 a particular size were an order of magnitude $(\pm 0.005 \text{ mm})$ less than the diameter difference (0.05

191 mm) between adjacent pin sizes. Pins of one size did vary in mass $(\pm 0.2{\text -}0.9 \text{ mg})$ with this

variation increasing with larger pins, but it usually represented a small component of the overall

mass. In the few cases where variation in pin mass represented >12.5% of the total weight, the

specimen was unmounted and weighed directly.

High humidity slowed analysis as the balances required longer to stabilize, but changes in mass

linked to variation in temperature and humidity were small. For example, the mean standard

error based on seven measurements of 20 specimens over a week was 1–2% of their mean.

Comparison of specimen weights between 2014 and 2018 further indicated that differences

- 200 between paired measurements were within \pm 3% of their average in 117 of 120 cases while the
- 201 others were within \pm 5%. A t-test demonstrated that mean mass increased by 0.3% ($p = 0.02$)
- over the interval, likely reflecting higher humidity when the second measurements were made.
-

Analysis further established that the loss of an appendage had a small impact on mass. For

205 example, a leg typically represented $1-2\%$ of the specimen's mass, while the tibia plus tarsus

- were around 0.5%. The loss of a major body segment (head, abdomen) had much larger impacts
- as they comprised about 12% and 50% respectively of the total mass.
-

Discussion

Because of its strong association with crucial biological traits, mass data is valuable in many

- ecological and evolutionary contexts. By assembling mass data for 3,161 species of Coleoptera,
- this study confirms that museum specimens are a valuable resource for constructing a mass
- registry. It further demonstrates that factors such as the loss of an appendage, variation in
- humidity, specimen age, and curatorial practices have small impacts on these measurements.
-

This study further demonstrates that mass variation among beetles has strong phylogenetic

217 I has study further demonstrated that these theories of the family, subfamily, and generic levels. As a $\frac{1}{\sqrt{2}}$

219 consequence, the analysis of a single or a few individuals of a species provides a good estimate

of its mass. Prior studies have demonstrated that adult body size can be impacted by diverse

- 221 environmental factors and that the extent of such variation differs among species (Emlen $\&$
- Allen, 2003; Chown & Gaston, 2010; Tseng et al., 2018). These impacts can even shift the
- relationship between morphometric measurements (e.g., body length) and biomass (Gouws,
- Gaston & Chown, 2011). Given the millions of insect species, it is not feasible to investigate
- such impacts on a species-by-species basis, but it is unnecessary in most contexts because they cause minor modulations in body mass as > 90% of variation resides at higher taxonomic levels
- 227 (Chown & Gaston, 2010). Importantly, the accuracy of mass estimates derived from generic
- assignments is similar to those resulting from the standard approach to mass estimation: the use
- of a power equation to estimate mass from body length (Rogers, Hinds & Buschbom, 1976).
- While our data set is larger and spans a greater range of mass, our residual SE (0.62) was less
- than that resulting from the use of a power equation (Table 1.c). In fact, even when our analysis
- targeted genera with the most variation in mass among their component species, the residual SE
- 233 (0.69) was similar to that (0.66) reported with the use of a power equation (Rogers, Hinds $\&$ Buschbom, 1976) (Table 1.d). While this residual translates into an average two-fold difference
- in mass from the predicted value, it indicates that a generic assignment can generate mass
- estimates with a precision similar to those based on estimates from direct length measurements.
- In short, mass values for a few species in each genus allow the estimation of mass for congeneric
- taxa. Although our results only document this fact for Coleoptera, similar relationships
- undoubtedly extend to other groups, as strong phylogenetic signal in body size occurs in many

arthropod lineages (Rainford, Hofreiter & Mayhew, 2016). Understanding the extent of

phylogenetic constraint in arthropods could greatly speed the development of a functional mass

- registry by allowing analysis to focus on groups where size variation is most pronounced and to
- use proxy measures in those where it is not.
-

 By delivering information on body mass for about 0.2% of all described insect species, the present study indicates that it is feasible to construct a mass registry for all insects. Furthermore, it reveals shortcuts to develop this registry. Specifically, the strong phylogenetic constraints on mass indicate that early efforts should focus on gaining coverage for higher taxonomic categories – every insect family and subfamily. Work should then extend to every genus and in time to every species. Because this effort will generate a substantial volume of data, it needs a home and 251 the BOLD platform (Ratnasingham & Hebert, 2007) is well-suited to meet this need. Although specimens with mass data need not possess barcode records, the inclusion of sequence information will maximize the utility of these records for metabarcoding and eDNA analysis. Moreover, the barcode records will ensure that specimens in the mass registry are properly identified, one of the key problems confronting any large-scale repository of biological collaterals. To demonstrate its capacity, the current records are deposited in a dataset on BOLD 257 (DS-MASSCOL; dx.doi.org/10.XYZZ/DS-MASSCOL **NOTE** to editors and reviewers: a DOI has been requested for the dataset, but it was not yet available at the time of submission]) that couples barcode records with mass information on each specimen examined in this study. Because DS-MASSCOL is a dynamic dataset where BIN assignments may shift and where specimens that currently lack a genus or species assignment may gain one, a supplemental file (Data S2) provides a snapshot at the time of submission.

Conclusions

Aside from its value on providing a basis for extending understanding of the evolution of body

mass, comprehensive body mass data on insect species is needed for ecological modeling. By

confirming that variable curatorial and preservation practices have little impact on body mass,

the present study establishes that museum specimens provide a resource for the rapid assembly

- of mass data. Employing this approach, the present study assembled mass data for 3,161 species
- of Coleoptera, nearly 1% of all known species in this order. Moreover, because of the strong
- phylogenetic constraints on body size, the current records enable accurate mass estimation (+/- 20%) for nearly all Canadian beetles. The extension of this approach to other arthropod groups
- and other geographic regions would facilitate the assembly of a mass registry for all insects.
- 276 Incorporation of the resultant mass value \equiv for each BIN into the parameters on BOLD will
- ensure easy access to these data.
-
-

References

PeerJ

 Gowing G, Recher HF. 1984. Length-weight relationships for invertebrates from forests in south-eastern New South Wales. *Australian Journal of Ecology* 9: 5-8 DOI: 10.1111/j.1442- 9993.1984.tb01612.x Greve ME, Hager J, Weisser WW, Schall P, Gossner MM, Feldhaar H. 2018. Effect of forest management on temperate ant communities. *Ecosphere* 9: e02303 DOI: 10.1002/ecs2.2303 Hebert PDN, Cywinska A, Ball SL, deWaard JR. 2003. Biological identifications through DNA barcodes. *Proceedings of the Royal Society of London. Series B: Biological Sciences* 270: 313- 321 DOI: 10.1098/rspb.2002.2218 Hendrich L, Morinière J, Haszprunar G, Hebert PDN, Hausmann A, Köhler F, Balke M. 2015. A comprehensive DNA barcode database for Central European beetles with a focus on Germany: adding more than 3,500 identified species to BOLD. *Molecular Ecology Resources* 15: 795-818 DOI: 10.1111/1755-0998.12354 Johnston TA, Cunjak RA. 1999. Dry mass-length relationships for benthic insects: a review with new data from Catamaran Brook, New Brunswick, Canada. *Freshwater Biology* 41: 653-674 DOI: 10.1046/j.1365-2427.1999.00400.x Jones KE, Bielby J, Cardillo M, Fritz SA, O'Dell J, Orme CDL, et al. (2009) PanTHERIA: a species-level database of life history, ecology, and geography of extant and recently extinct mammals. *Ecology* 90: 2648-2648 DOI: 10.1890/08-1494.1 Knapp M. 2012. Preservative fluid and storage conditions alter body mass estimation in a terrestrial insect. *Entomologia Experimentalis et Applicata* 143: 185-190 DOI: 10.1111/j.1570- 7458.2012.01247.x Novotny V, Kindlmann P. 1996. Distribution of body sizes in arthropod taxa and communities. *Oikos* 75: 75-82 DOI: 10.2307/3546323 Pentinsaari M, Hebert PDN, Mutanen M. 2014. Barcoding beetles: a regional survey of 1872 species reveals high identification success and unusually deep interspecific divergences. *PLoS One* 9: e108651 DOI: 10.1371/journal.pone.0108651 Pentinsaari M, Anderson R, Borowiec L, Bouchard P, Brunke A, Douglas H, Smith ABT, Hebert PDN. 2019. DNA barcodes reveal 63 overlooked species of Canadian beetles (Insecta, Coleoptera). *ZooKeys* 894: 53-150 DOI: 10.3897/zookeys.894.37862

PeerJ

PeerJ

Figure 1

Distribution of the log-transformed body masses for 3,161 BINs of Canadian Coleoptera.

Figure 2

Distribution of the variance in body mass for Coleoptera genera.

Interquartile range measures the difference between the upper and lower quartiles and can be converted to fold-difference or used to estimate the typical deviation from the median (e.g., IQR 0.2 = 1.6-fold difference between quartiles $\approx +1/23$ % from median; IQR 0.4 = 2.5fold difference between quartiles $≈ +/-$ 57% from median).

Table 1(on next page)

Output of one-way and nested analyses of variance.

1 **Table 1.** *Output of one-way and nested analyses of variance.*

2