Phenotypic Correlates of Genome Size in Lepidoptera

Author: Miller, William E.

PHENOTYPIC CORRELATES OF GENOME SIZE IN LEPIDOPTERA

WILLIAM E. MILLER

Department of Entomology, University of Minnesota, 1980 Folwell Ave., St. Paul, Minnesota 55108; mille014@umn.edu.

ABSTRACT. Body size, developmental rate, and metabolic rate were examined as potential phenotypic correlates of genome size using all 51 named species of Lepidoptera with recorded genome sizes. Genome sizes ranged 0.29–1.94 pg. Because no direct comparative measures were available, surrogates were used: wingspan for body size, voltinism for developmental rate, and grades of adult diel activity and adult feeding for metabolic rate. Analyses consisted of plotting genome size on surrogate values and fitting least-squares trend lines. At the order level, empirical - correlates of genome size with wingspan and voltinism were essentially null or flat. At family and subfamily levels, striking associations were found for wingspan and voltinism, but not for surrogates of metabolic rate, for which data were limited and surrogate validity uncertain. Geometrid species, nearly all of which were ennomiines, showed a sharp positive association between genome size and wingspan, Arctiidae a sharp negative association, and Noctuidae a flat association. Geometridae showed a sharp negative association between genome size and voltinism, Arctiide a sharp positive association, and Noctuidae a mild negative association. No phenotypic correlates of lepidopteran genome size were previously known. Suppositions were confirmed that phenotypic correlates are most likely to be detected at family and lower taxonomic levels.

Additional key words: Body size, developmental rate, voltinism, metabolic rate, diel activity, adult feeding.

The genome is the entirety of an organism's hereditary information. It consists of the amount of DNA or total number of DNA base pairs in one genome copy, typically as observed in the nucleus of a male sperm cell. Size is a key genome element and genome size a fundamental species trait. It is usually measured as mass or weight in picograms (pg, trillionths or 10–12 gram), sometimes as number of base pairs in megabases (Mb), with 978 Mb equal to 1 pg. Genome size is also known as C-value and as haploid genome number. Genome basics are discussed by Benfey & Protopapas (2005).

Most of the reported lepidopteran genome sizes are due to Gregory & Hebert (2003). They provided more than five times the previously known number, bringing to 51 the named species in 11 moth and 3 butterfly families with recorded genome sizes (Appendix 1), as well as several unnamed species. This dataset ranks third after Coleoptera and Diptera in number of genome size records for holometabolous insect orders (Gregory 2012). Genome sizes of the 51 named species, all used in this study, ranged 0.29–1.94 pg, averaging 0.67 pg.

In sexual multicellular organisms, or eukaryotes, genome size is typically assessed using densitometric measurements of Feulgen-stained nuclei and either computerized image analysis or flow cytometry. Measurement is discussed by Hardie et al. (2002) and briefly summarized as follows. The tightly bundled spermatozoa are first separated to obtain individual nuclei of eupyrene sperm. Spermathecae of mated females can also be dissected to recover spermatozoa. Next, nuclei are Feulgen stained, which consists of fixation, hydrolysis, and staining in Schiff reagent, followed by rinsing and drying. Integrated optical densities of stained nuclei are then examined using image-analysis software with a special camera. Finally, results are compared against standards such as *Drosophila melanogaster* for conversion to absolute DNA content. Genome size is a precise quantity with coefficients of variation generally less than 5 percent as computed from data in Gregory & Hebert (2003) for sample sizes of 3 or more. This variation is attributed to measurement error.

Prior studies with animals have shown that genome size is neither proportional to number of genes nor closely related to organismal complexity. Its relevance must be revealed by empirical investigation. Phenotypic correlates help to illuminate genome biology and evolution. Correlates have been reported at cell and organism levels, and include cell size, cell division rate, body size, metabolic rate, and developmental rate (Gregory 2002). Three common correlates in animals are body size, developmental rate, and metabolic rate (Gregory & Hebert 2003), and the present report explores the association of these three variables with lepidopteran genome size. Surrogates for the variables were necessarily used because direct comparative measurements were not possible. As yet, no phenotypic correlates have been reported in Lepidoptera.

METHODS AND RESULTS

The intent in all analyses was to discover associations between genome size and the surrogates. All analyses were relational, and to avoid abetting error, genome size was plotted on surrogates rather than on back-conversions to original variables. Straight lines of the form $y_1 = a + b\, y_2$ were fitted through data points by minimizing the sum of squares of differences between lines and points, as in typical regression. However,
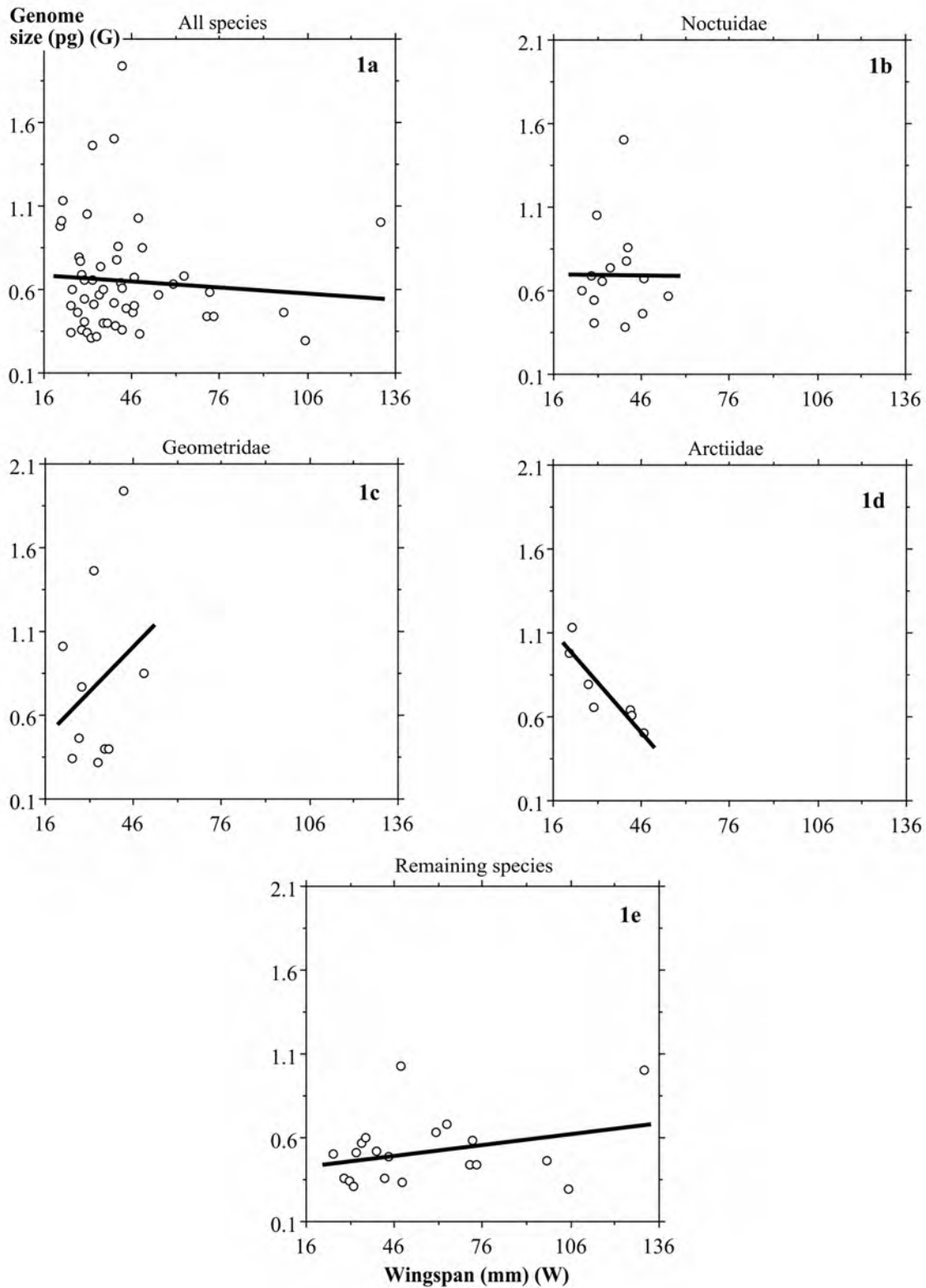
FIGS. 1a–1e. Empirical associations between genome size (pg) (G) and wingspan (mm) (W). **1a,** all species, G = 0.71 - 0.0013 W, n=51. **1b,** Noctuidae, G = 0.72 - 0.00046 W, n=14. **1c,** Geometridae, G = 0.19 + 0.018 W, n=10. **1d,** Arctiidae, G = 1.40 - 0.019 W, n=7; and **1e,** remaining species omitting Noctuidae, Geometridae, and Arctiidae, G = 0.40 - 0.0022 W, n=20.

except where noted, no statistical regression or correlation is implied here, the lines being used only for description. The terms 'null' and 'flat' describe lines whose slope coefficients, b in the above equation, appear to be little different than zero. Study results were not routinely evaluated statistically because of small sample sizes and uncertain conformity to test assumptions. For simplicity, lepidopteran nomenclature and classification here follow Gregory & Hebert (2003). Also, only the first of two nearly identical genome sizes for *Bombyx mori* was used.

Body size. The most precise measure of body size is mass or weight, but such information is scarce for all life stages of Lepidoptera. The comparative surrogate used was wingspan, which has been recorded and is retrievable online for virtually all lepidopteran species. Midrange or median values were mostly used here (Appendix), except where only one value was reported. The assembled wingspans ranged 21.5–131 mm and averaged 43.3 (n=51). From studies of Tortricidae and Sphingidae, which encompass small to large body sizes, body mass is closely associated with wing measure at the family level (Miller 1977, 1997).

Genome sizes plotted on wingspans are shown in Figs. 1a–1e, all on the same scale for easy comparison. The trend line of the all-species plot is essentially flat (slopes not significantly different than 0), suggesting no association at order level between genome size and wingspan (Fig. 1a, n=51). Breakout of plots for the three families with the most genome-size records (Figs. 1b–1d) disclosed differing associations, most in the lower ranges of both genome size and wingspan. In Noctuidae there was a null association (Fig. 1b) (genome size 0.38–1.50 pg, averaging 0.71, wingspan 25.5–55 mm, averaging 0.37.5, n=14); in Geometridae there was a sharp positive association (Fig. 1c) (genome size 0.32–1.94 pg, averaging 0.80, wingspan 22–49.5 mm, averaging 33.5, n=10); and in Arctiidae, a sharp negative association (Fig. 1d) (genome size 0.50–1.13 pg, averaging 0.76, wingspan 21.5–47 mm, averaging 33.4, n=7). With data of these three families removed, the remaining-species plot showed a flat association like the all-species plot (Fig. 1e) (genome size 0.29–1.03 pg, averaging 0.52, wingspan 25–131 mm, averaging 53, n=20) (n=1–4 per family).

The genome size-wingspan associations evident in Geometridae and Arctiidae (Figs. 1c, 1d) are striking, and their steep slopes despite small samples suggest strong relations. They validate the efficacy of wingspan as a body-size surrogate, and indicate likely phenotypic correlates. Gregory & Hebert (2003) noted that important differences in genome size may occur at the subfamily level. The 14 named species of Noctuidae are thought to represent 12 subfamilies, which suggests high heterogeneity, and may help to explain the apparent null family relation (Fig. 1b). Subfamily level associations in the overall dataset (n=1–9 per subfamily) were not routinely investigated. However, 9 of the 10 geometrids studied are usually considered to be ennomiines, and all 7 arctiids to be arctiines. Genome sizes of more lepidopterans are needed to confirm and expand the foregoing associations.

Developmental rate. Voltinism was taken as a comparative surrogate for developmental rate or speed. To illustrate, bivoltinism, the completion of two complete generations in a single growth season, was assumed to indicate development twice as fast as univoltinism. Voltinism information was found online for 49 of the 51 named species (Appendix). In analysis, one generation per year was coded as 1, two per year as 2, three per year as 3, and so on. For some species, voltinism depends on latitude (*egle, tesselaris, intermedia, grata, pallescens* (=*furcilla*), *gibbosa, excaecatus*) with one annual generation in the North and two or more southward. To better represent full voltinism capability and comparability with univoltine obligates such as *americana* and *disstria*, maximum reported voltinism, regardless of latitude, was used.

Genome size was plotted on voltinism using the same scale for all taxonomic categories (Figs. 2a–2e), and results are shown in the same taxonomic sequence as in Figs. 1a–1e. The all-species trend line seemed negative (Fig. 2a) (genome size 0.29–1.94 pg, averaging 0.67, voltinism 1.0–6.0 averaging 2.3, n=49). The trend in Noctuidae (Fig. 2b), again representing a number of subfamilies, was mildly negative (genome size 0.38–1.50 pg, averaging 0.71, voltinism 1.0–6.0, averaging 2.9, n=13n). A clearly positive trend appeared in Arctiidae (Fig. 2d) (genome size 0.5–1.13 pg, averaging 0.76, voltinism 2.0–3.0, averaging 2.3, n=7), although this result appears to be influenced by one species. In Geometridae, the trend was sharply negative, and as mentioned earlier, all but one geometrid data point likely represented the one subfamily Ennomiinae (Fig. 2c) (genome size 0.32–1.94 pg, averaging 0.84, voltinism 1.0–2.0, averaging 1.5, n=9), thus giving further credence to the idea that investigations below family level are more likely to reveal correlates. A mildly positive trend appeared in species remaining after removal of the foregoing three families (Fig. 2e) (genome size 0.29–1.0 pg, averaging 0.54, voltinism 1.0–5.9, averaging 2.4, n=18).

Negative associations of genome size with voltinism as seen in Noctuidae and Geometridae (Figs. 2b, 2c) are typical of other invertebrate groups with approximately uniform complexity (Gregory 2002). Also, the
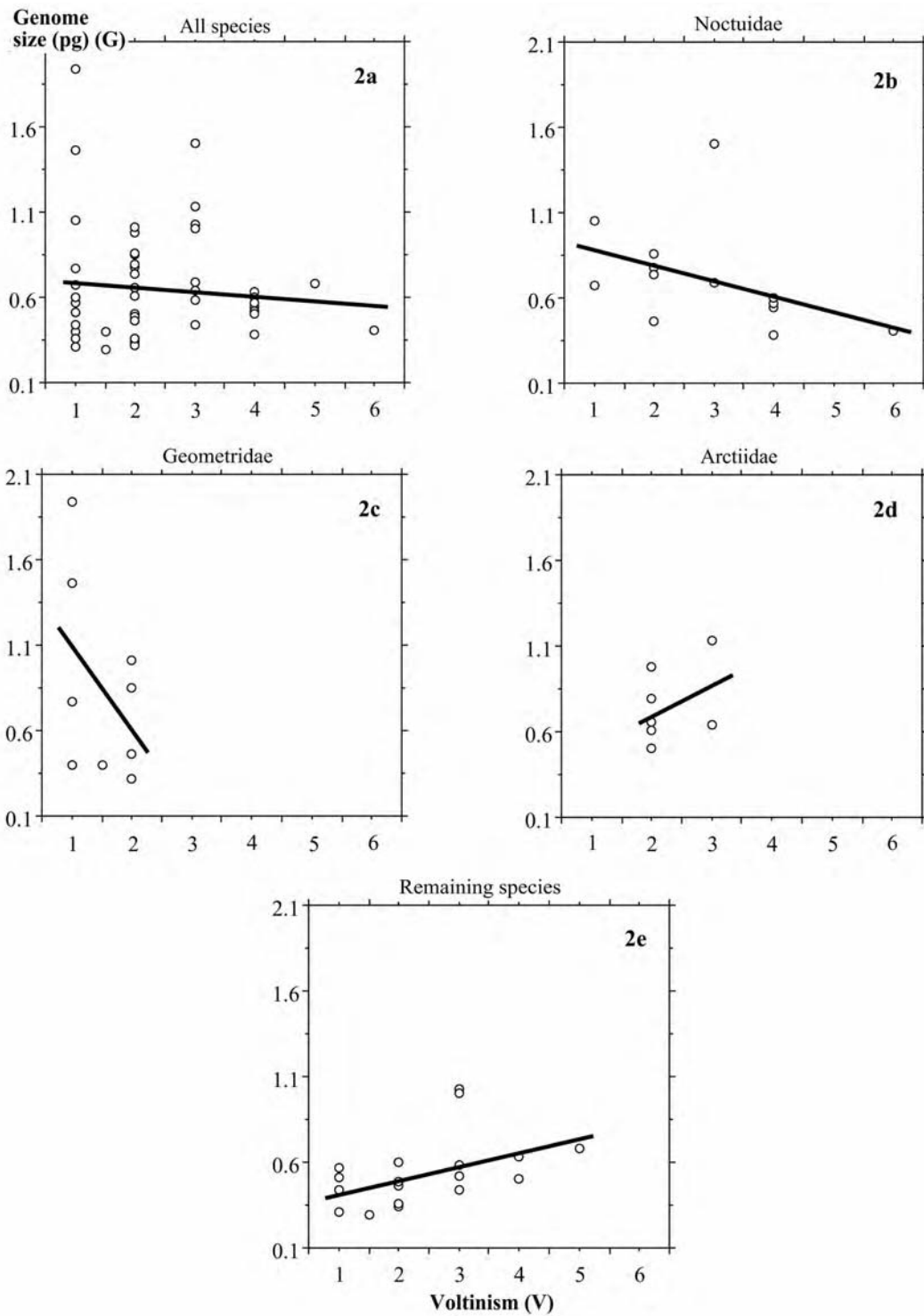
FIGS. 2 a–e. Empirical associations between genome size (pg) (G) and voltinism (V). **2a,** all species, G = 0.73 - 0.028 V, n=49; **2b,** Noctuidae, G = 0.98 - 0.0.090 V, n=13; **2c,** Geometridae, G =1.57- 0.48 V, n=9; **2d,** Arctiidae, G = 0.35 - 0.18 V, n=7; and **2e,** remaining species omitting Noctuidae, Geometridae, and Arctiidae, G = 0.40 + 0.0022 V, n=18.

Geometridae-Ennomiinae results suggest that fastest development occurs in species with small genomes.

Metabolic rate. Identifying a surrogate for metabolic rate proved problematic. Gregory & Hebert (2003) suggested that powered flight might offer clues to associations between metabolism and genome size. Powered flight may be diurnal, nocturnal, or mixed. In a meticulous study of 24-h adult diel activity across many lepidopteran taxa, Fullard & Napoleone (2001) calculated diel nocturnality percentage for more than 80 species. They devised a four-class system for characterizing diel activity, with classes defined by nocturnality percentage as follows: 0–10, exclusively diurnal; 10–50, primarily diurnal; 50–90, nocturnal; and 90–100, exclusively nocturnal. Genome sizes were available for 15 of these species (Appendix). When these genome sizes were plotted on corresponding diel activity class values (Fig. 3), the resulting association was flat, but there were no more than three data points for any family. To enlarge this sample, diel activity class for 19 more species was approximated (Appendix) with the help of online sources (Fig. 3). After a Student t-test showed that the two arrays did not likely differ (t = 1.26, P = 0.22), they were combined, enlarging the all-species sample to 34n (Appendix). Again, a flat slope resulted (Fig. 3). At the family level, associations between genome size and diel activity class were likewise essentially null, with slope coefficients 0.054 for Noctuidae (n=7), 0.0050 for Geometridae (n=5), and 0.027 for Arctiidae (n=7) (graphs not shown).

The potential of adult diel activity as a surrogate hinges on accurate metabolic characterization of diurnality and nocturnality. The characterization here may be flawed. Questions arise whether metabolic rate increases or decreases along the spectrum between nocturnality and diurnality, and where on the spectrum mixed nocturnality-diurnality belongs. It is also unclear whether species capable of evasive nocturnal flight triggered by auditory bat detection (*bilineata, gibbosa, dispar, grata, pallescens* and others) (Fullard & Napoleone 2001) differ in metabolism from species lacking this capability. Another prominent factor of unknown influence on metabolism is female flightlessness (*pomonaria, subsignaria, dispar*). Some of these factors may explain the seeming failure of diel activity class as a surrogate for metabolic rate, but the small numbers of observations per family (n=5–7) may also be involved. Larger samples might tell a different story.

Amount of food consumed by adults also seemed a possible surrogate for metabolic rate. Species of known genome size were classified with the help of online sources into three earlier established adult feeding classes: non-feeders with vestigial proboscises, adults capable of limited feeding, and adults feeding heavily (Miller 1996). For analysis here these were coded 1, 2, and 3, respectively (Appendix). The resulting all-species plot of genome size on adult feeding class produced an essentially null slope of 0.036 (n=28) (graph not shown). Geometridae, Arctiidae, and Noctuidae plots had five or
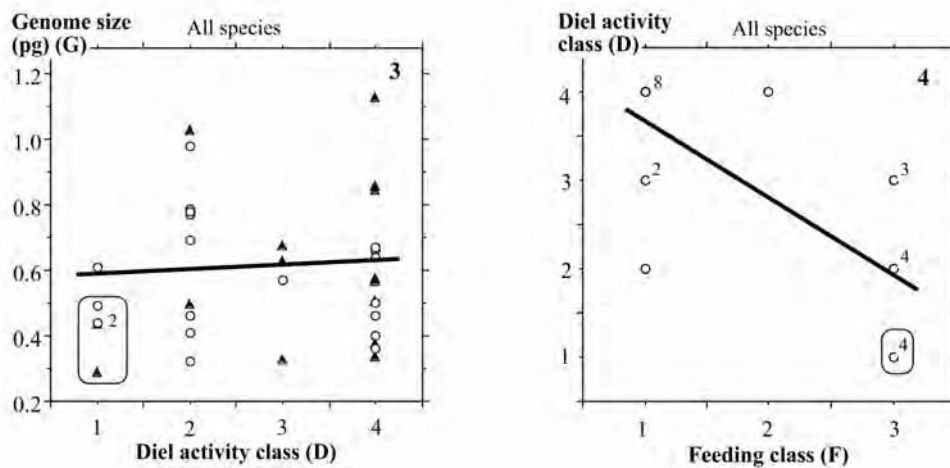

FIG. 3. Empirical association between genome size and diel activity class. Filled triangles represent values based on Fullard & Napoleone (2001); open circles, values approximated here using online sources; and boxed values, species in butterfly families. Slope coefficients were: for all species 0.016, n=34; for Noctuidae, 0.14, n=7; for Geometridae, 0.054, n=5; for Arctiidae 0.027, n=7; and for remaining species omitting Noctuidae, Geometridae, and Arctiidae, 0.00042, n=15.

FIG. 4. Relation between diel activity class (D) and adult feeding class (F). D = 4.55 - 0.364 F, n=23. Boxed values represent species in butterfly families. Statistical testing explained in text.

fewer points each (Appendix), too few for reliable conclusions.

The four species in the study from butterfly families are boxed in Figs. 3 and 4.

One association emerged that could have future usefulness in identifying a surrogate, namely that between adult diel activity class and adult feeding class (Fig. 4). This association was highly significant statistically (Olmstead-Tukey corner test, S = 20, P = 0.001, n=23) (Sokal & Rohlf 1995). It demonstrates how increasing diurnality is accompanied by increasing adult feeding. Either variable might eventually have surrogate value if more data accumulates that enables either one to predict the other toward augmenting sample size. Using adult feeding class here to estimate more diel activity class values was considered but not carried out because only six new values would have been created, too few to make a difference in any family.

DISCUSSION

Body size as represented by wingspan proved a likely phenotypic correlate, notably in Geometridae where its association with genome size was clearly positive (Fig. 1c), and in Arctiidae where it was clearly negative (Fig. 1d). It may be noteworthy that body size and its surrogate, like genome size, are measured in units of mass, and that body size, again like genome size, is a precise end-value.

Developmental rate or speed, as represented by voltinism, also seemed a likely phenotypic correlate, especially in Geometridae where its association with genome size was negative (Fig. 2c), and possibly in Noctuidae where the association also appeared to be negative (Fig. 2b). Developmental rate and voltinism, unlike body size and wingspan, were not assessed in the same units as genome size. Both are imprecise in the present context in that they are processes rather than end-points, and voltinism is of necessity a categorical rather than continuous variable. Also, voltinism may be influenced by variation in environmental factors, and, moreover, could relate directly to genome size apart from developmental rate. Metabolic rate as a potential correlate thus requires more research. In particular, a suitable surrogate or means of interspecific comparison is needed.

It is evident from this study that more can be learned about phenotypic correlates in Lepidoptera at family and lower taxonomic levels than at the order (all-species) level. This confirms a conclusion reached on other grounds by Gregory & Hebert (2003). The most useful future research in the quest for phenotypic correlates of lepidopteran genome size would be measurement of genome size in more species, and identification of new potential correlates and surrogates. Correlates might turn out to abound if the right investigative tools are deployed. Hopefully the mechanistic reasons for these correlates will be better understood in future studies.

LITERATURE CITED

BENFEY, P. N. & A. D. PROTOPAPAS. 2005. Essentials of genomics. Prentice-Hall. Upper Saddle River, N. J.

FULLARD, J. H. & N. NAPOLEONE. 2001. Diel flight periodicity and the evolution of auditory defenses in the Macrolepidoptera. Animal Behav. 62: 349–368.

GREGORY, T. R. 2002. Genome size and developmental complexity. Genetica. 115: 131–146.

GREGORY, T. R. 2012. Animal genome size database. http://wwwgenomesize.com.

GREGORY, T. R. & P. D. N. HEBERT. 2003. Genome size variation in lepidopteran insects. Canadian J. Zool. 81: 1399–1405.

HARDIE, D. C., T. R. GREGORY & P. D. N. HEBERT. 2002. From pixels to picograms: a beginner's guide to genome quantification by Feulgen image analysis densitometry. J. Histochem. Cytochem. 50:735–749. (www.ncbi.nih.gov/pubmed/12019291).

MILLER, W. E. 1977. Wing measure as a size index in Lepidoptera: the family Olethreutidae. Ann. Entomol. Soc. Amer. 70: 253–256.

MILLER, W. E. 1996. Population behavior and adult feeding capability in Lepidoptera Environ. Entomol. 25: 213–226.

MILLER, W. E. 1997. Body weight as related to wing measure in hawk-moths (Sphingidae). J. Lepid. Soc. 51: 91–92.

SOKAL, R. R. & F. J. ROHLF. 1995. Biometry, ed. 3. Freeman, New York. 887 pp.

Editor's Note

This manuscript was in the revision stage when Dr. William Miller passed away. Dr. Marc Epstein and I completed the revision to the best of our abilities. The work remains entirely a product of Dr. Miller's creative mind.

APPENDIX. Complete enumeration of data used in this study. Generic assignments of species follow Gregory & Hebert (2003). Dash signifies unavailable datum. Diel activity class values marked with ‡ are based on data in Fullard & Napoleone (2001); those without ‡ are approximations gleaned from online sources. Further explanation in text.

Taxon	Genome size (pg)	Wing Span (mm)	Voltinism	Diel Activity Class	Adult Feeding Class
Arctiidae					
Ctenucha virginica	0.50	47	2	2‡	3
Euchaetes egle	0.79	28	2	2	–
Grammia virgo	0.98	21.5	2	2	–
Halysidota tesselaris	0.61	42.5	2	3	3
Hyphantria cunea	0.66	30	2	4	1
Hypoprepia fucosa	1.13	22.5	3	4‡	–
Spilosoma virginica	0.64	42	3	4	–
Bombycidae					
Bombyx mori	0.52	40	4	–	1
Danaidae					
Danaus plexippus	0.29	105	1.5	1‡	3
Drepanidae					
Drepana arcuata	0.31	32	1	–	–
Drepana bilineata	0.34	30.5	2	4‡	–
Geometridae					
Ennomos subsignaria	0.40	37.5	1	4	–
Euchlaena irraria	1.94	42.5	1	–	–
Gueneria similaria	0.40	36	1.5	–	–
Lycia (Poecilopsis) pomonaria	1.46	32.5	1	–	–
Metanema determinata	1.01	22	2	–	2
Nepytia canosaria	0.77	28.5	1	2	–
Prochoerodes transversata	0.85	49.5	2	4‡	2
Scopula limboundata	0.46	27.5	2	2	–
Semiothisa bisignata	0.34	25	–	–	–
sospeta	0.32	34	2	2	–
Lasiocampidae					
Malacosoma americana	0.51	33	1	4‡	1
Malacosoma disstria	0.57	35	1	4‡	1
Lymantriidae					
Lymantria dispar	0.03	48	3	2‡	1

Appendix continued …

APPENDIX. (continued)

Taxon	Genome size (pg)	Wing Span (mm)	Voltinism	Diel Activity Class	Adult Feeding Class
Noctuidae					
Caenurgina crassiuscula	0.50	140	3	–	–
Calophasia lunula	0.69	29	3	2	3
Cerma cerintha	1.05	30.5	1	–	–
Condica videns	0.54	30	4	–	–
Cucullia intermedia	0.46	46.5	2	–	–
Eudryas grata	0.38	40.5	4	4‡	1
Heliothus virescens	0.41	30	6	2	3
Lacinipolia renigera	0.60	25.5	4	–	–
Macronoctua onusta	0.67	47	1	4	–
Noctua pronuba	0.57	55	4	3	3
Panthea pallescens	0.86	41.5	2	4‡	–
Papaipema nepheleptena	0.66	32.5	–	–	–
Raphia frater	0.74	35.5	2	–	–
Xestia (Amathes) c-nigrum	0.78	41	2	2	3
Notodontidae					
Gluphisia septentrionis	0.36	29	1	4	1
Heterocampa bilineata	0.60	36	1	–	1
Nadata gibbosa	0.33	48.5	2	3‡	3
Peridea basitriens	0.36	42.5	2	–	1
Papilionidae					
Papilio canadensis	0.44	71.5	1	1	3
Papilio glaucus	0.44	74	3	1‡	3
Pyralidae					
Galleria melonella	0.50	25	4	4	1
Saturniidae					
Antheraea pernyi	1.00	131	3	–	1
Satyridae					
Bicyclus anynana	0.49	44	2	1	3
Sphingidae					
Pachysphinx modesta	0.46	98	2	4	1
Paonias excaecatus	0.58	72.5	3	4‡	1
Paonias myops	0.63	60	4	3‡	1
Smerinthus jamaicensis	0.68	64	5	3‡	1