

- Chapter Six -

Genome size evolution in terrestrial arthropods

- Chapter Six -
Genome size evolution in terrestrial arthropods¹

Abstract

Any comprehensive discussion of animal evolution must necessarily include information on insects, but unfortunately this most speciose of all animal groups has been relatively poorly studied from the perspective of genome size. Roughly 90% of the existing insect C-value estimates come from just four orders (Coleoptera, Diptera, Hemiptera, and Orthoptera), and only 10 orders have previously been studied at all. In an effort to correct this major taxonomic omission, this chapter adds 230 new insect C-value estimates to the dataset. In combination with previously published values, this allows at least a descriptive analysis of the major patterns in insect genome sizes. In particular, the notion of developmental complexity introduced in the discussion of amphibian genome sizes (Chapter 4) appears to apply to the insects as well. New C-value estimates are also provided for over 110 species of spiders and several myriapods, the first so far reported for either group.

¹

This chapter consists largely of material not yet published, but contains information taken from Gregory (2002b,d).

“It seems therefore that a taste for collecting beetles
is some indication of future success in life.”
- Charles Darwin

Introduction

The fact that most animals are insects means that no general discussion of metazoan evolution can be complete without reference to the Insecta. Put in terms of the C-value enigma, most animal genomes reside within insect cells, so no discussion that ignores this group can be considered comprehensive. Despite this, most of the current information on animal genome sizes continues to stem from studies of vertebrates. To be sure, analyses of patterns in vertebrates are useful in establishing the general framework for genome size evolution in animals (e.g., Chapters 3 and 4). However, the applicability of these principles to animals at large remains an unresolved issue, primarily because so few data are available for the most speciose group in the kingdom.

This chapter provides a review of available insect genome size data, and also includes a large number of new estimates which effectively double the size of the database and expand its coverage greatly. Yet, even with this enlarged dataset, the major patterns within and among groups of insects remain somewhat ambiguous. This chapter also provides the first large survey of genome sizes from North American spiders and the first C-value estimates from myriapods, two groups of terrestrial arthropods with no current representation in the animal genome size dataset. While still too superficial to reveal any obvious patterns, the current dataset does provide a much-needed introduction to the study of spider and myriapod genome size evolution, and outlines some potentially fruitful areas of future research in this area. A small number of arachnids other than spiders are also discussed, although no broad sampling effort has yet been attempted for

these taxa. In this regard, the present chapter should be considered an introduction to arthropod genome size study, not a conclusion.

Summary of the dataset(s)

Insects: previously published data

Appendix 11 contains a total of about 290 insect species representing 45 families from 10 of the 32 (or so) orders. Although this is a larger dataset than is currently available for birds (Appendix 7) and only slightly smaller than that for reptiles (Appendix 9), in relative terms it covers at best about 0.0004% of the insect class. Moreover, 90% of the currently available estimates are from just four orders: Coleoptera, Diptera, Hemiptera (almost all from the Suborder Homoptera), and Orthoptera (Table 9.1).

Insects: new genome size estimates

In an effort to expand the insect genome size dataset to the point where a useful comparative commentary could be offered, the present study included estimates for an additional 230 species from 73 families covering 18 orders (Appendix 12). The Coleoptera served as the main focal group in the present survey, but the Lepidoptera, Hemiptera (especially the Suborder Heteroptera), and Orthoptera were also well represented. Smaller numbers of specimens from as many additional orders as could be collected were also included in the present study (Table 9.1).

Most of the insects measured here were collected from sites in Southern Ontario (Appendix 12). The basic methods used for slide preparation were those outlined in Chapter 5, and some additional comments on methodology are given below as appropriate. The primary tissues analyzed were sperm and haemocytes, and in most cases 25-30 nuclei were measured per specimen (however, more or fewer nuclei were measured

in some cases, as dictated by the quality of the preparations). As seen in Appendix 12, error among conspecifics was very low. The majority of species identifications were kindly performed by Marvin Gunderman (McMaster University; Coleoptera), Paul Hebert (University of Guelph; Lepidoptera), and Steve Marshall (University of Guelph; Hemiptera, Orthopteran, and miscellaneous others).

Spiders

To date, only one arachnid C-value has been published, that of the tick *Amblyomma americanum* at 1.08pg (Palmer et al. 1994). Ellen Rasch (personal communication) has estimated the C-value of the brown recluse spider, *Loxosceles reclusa*, at 1.15pg, but to my knowledge no other spiders have yet been analyzed for genome size. When one considers the diversity and abundance of spiders – more than 35,000 species in about 100 families described so far, or about 40% more than teleost fishes – this omission becomes striking indeed.

All spiders are members of the arachnid Order Araneae, which is divided into two suborders. The Suborder Mesothelae represents a small group of primitive spiders, as evidenced by their retention of a clearly segmented abdomen. The more familiar spiders are grouped into the Suborder Opisthothelae, which is itself divided into two Infraorders. The Infraorder Mygalomorphae (formerly known as the Orthognatha because of the arrangement of their chelicerae) is comprised of all the relatively large ground-dwelling spiders commonly referred to as “tarantulas”². The larger Infraorder, the Araneomorphae

2

The name “tarantula” is derived from the Italian town of Taranto, and is properly reserved for the Mediterranean tarantula *Lycosa tarentula* and its close relatives. These are wolf spiders in the Family Lycosidae, and are members of the Infraorder Araneomorphae, not the Mygalomorphae (Foelix 1996).

(formerly Labidognatha), contains all the remaining spider families (and 90% of all spider species). The present dataset includes estimates for 114 species of spiders from 19 different families, all from the Infraorder Araneomorphae.

Spiders for the present survey were collected from various locations, mostly in southern Ontario. Haemolymph samples were prepared, stained, and measured as outlined in Chapter 5. Species identifications were kindly performed by David Shorthouse (University of Alberta), with some additional specimens (*Latrodectus* spp.) generously provided by Maydianne Andrade (University of Toronto).

In total, more than 15,000 nuclei were measured from roughly 500 individual animals. In most cases, 30 nuclei were measured per individual, although as few as 10 were analyzed with some of the most difficult preparations. However, since at least two individuals were available for the majority of species, most of the estimates were based on a minimum of 30 nuclei. CVs for nuclei from individual spiders averaged about 8%, which is slightly higher than the recommended level for vertebrate blood (< 6%), but still within the range suggested for invertebrate tissues by Hardie et al. (2002).

Chicken erythrocytes were used as the standard in this study because no well-established spider standards are yet available. It may therefore be necessary to adjust these values in the future if chicken erythrocyte nuclei are not found to share the DNA compaction level and staining characteristics of spider haemocyte nuclei, but even so the absolute error should be relatively minor and would not affect the overall trends discussed here.

Variation among conspecifics was minor, and could be readily attributed to measurement error. This was true even when the specimens were collected from different

populations around southern Ontario. In one particularly notable example, the three specimens of *Argiope aurantia* analyzed gave very consistent C-value estimates, despite having been collected from locales as disparate as Guelph, ON and St. Andrews, NB (CV < 4%). The entire spider dataset is given as Appendix 13, and includes information on the source locations and samples sizes for each species.

Other arachnids

In addition to the spiders, several specimens of the two-spotted spider mite, *Tetranychus urticae* (Order Acari, Family Tetranychidae), were prepared by the “blind mincing” technique (Chapter 5) and analyzed by measuring somatic nuclei. This species is of particular economic interest because it is a worldwide agricultural pest of more than 180 species of vegetables, food and fiber crops, fruit trees, ornamentals, and greenhouse plants (Osborne et al. 1999). Cytogenetically, this tiny chelicerate (adults ~ 0.5mm long) is of interest because it displays a haplo-diploid system of sex determination and a 3:1 female-biased sex ratio (Osborne et al. 1999). Both male (1C) and female (2C) somatic cells were measured in the present study. An unidentified red freshwater mite was also measured in the present study. Haemolymph preparations were made for several species of harvestmen (Order Opiliones) as well, but these did not stain well and the species were never identified.

Myriapods

Prior to the present study, no data were available on the genome sizes of centipedes or millipedes. As part of the present survey, 14 species of myriapods (3 centipedes, 11 millipedes) were assayed for variation in C-value (listed in Appendix 20 along with other miscellaneous invertebrates). The specimens were collected in Southern

Ontario and prepared as haemolymph slides as outlined in Chapter 5. A chicken erythrocyte standard was also used in this case, since no reliable myriapod standards are yet available.

Patterns of variation in insects

Although the insect genome size dataset has been substantially expanded in the present study, most groups of insects remain poorly sampled. As a result, patterns within orders can only be examined in any detail in a few cases. In the following sections the currently available insect C-value data are summarized on an (alphabetical) order-by-order basis³. Some potential trends within orders are discussed first, followed by a brief summary of additional patterns of DNA content modulation in insects. Finally, some broader comparisons among orders are provided at the end of the chapter.

Blattaria

Cockroaches are best known as pests in human dwellings, although only a very small percentage of species in the Order Blattaria actually warrant this status. In terms of genome size, however, these are the species that have been studied to date. The current list includes the American cockroach, *Periplaneta americana*, at 2.7pg, the German cockroach, *Blattella germanica*, at 2pg (but closer to 3pg in the present study), the Oriental cockroach, *Blatta orientalis*, at 3pg, and the giant roach, *Blaberus fuscus*, at 3.4pg. Despite the limited number of species, this does cover three different families, and

3

A note on taxonomy: Springtails (Class/Order Collembola) are usually considered hexapods, but not insects. Silverfishes (Order Thysanura) are insects, but are members of the wingless Subclass Dicondylia. All other insects discussed here are members of the Subclass Pterygota. The single Order Hemiptera is used here, including the Suborders Heteroptera and Homoptera, although this classification remains somewhat controversial as well.

it is worth noting that in all cases the genomes are relatively large in these insects. The C-values of two additional species not previously analyzed were estimated in the present study, and confirm this trend for large genome sizes among cockroaches. These include the wood roach, *Parcoblatta* sp., at 3.8pg and the Madagascan hissing cockroach, *Gromphadorhina portentosa*, at 3.4pg.

Coleoptera

Unfortunately, the “inordinate fondness for beetles” famously attributed to the Creator by J.B.S. Haldane does not appear to be shared by genome evolutionists. Despite their enormous diversity, beetles have been the subject of only a few genome size studies and are so far represented in the genome size dataset by less than 120 species from only six families (in North America alone, there are about 30,000 species in 111 beetle families). In total, this represents about 0.0003% of the roughly 350,000 species of coleopterans known throughout the world. The best (or rather, least poorly) studied families include the Chrysomelidae (leaf beetles) and the Tenebrionidae (darkling beetles), for which roughly 50 estimates each are available. The remaining dataset includes six larder beetles (Family Dermestidae), four weevils (Family Curculionidae), two diving beetles (Dytiscidae), and one ground beetle (Family Carabidae).

As a first step towards rectifying this neglect of the most diverse of all animal groups, the present study included estimates from as many families as could be collected (in one season by a non-expert) in southern Ontario. As a result of this relatively mild effort, genome sizes are now available for at least one representative of 25 different families (105 new estimates in total). Clearly, the limited sample of beetle genome sizes is not the product of insufficient material for study.

The patterns of genome size variation among the families of beetles studied to date is depicted in Figure 6.1. Overall, available beetle C-values range from 0.14pg in the false ant-like flower beetle *Pedilus* sp. (Family Pedilidae) to 2.2pg in an unidentified rove beetle (Family Staphylinidae)⁴. This range is somewhat misleading, however, since only eight of the more than 220 species studied so far display genomes larger than 1.5pg (Appendices 11 and 12), and the mean for the entire beetle dataset is only 0.61pg \pm 0.03. Moreover, only three families are known to contain at least two members with C-values larger than 1pg: the Chrysomelidae, Dermestidae, and Dytiscidae (Appendices 11 and 12). Similarly, only three families appear to have mean genome sizes larger than 1pg: Dermestidae, Dytiscidae, and Staphylinidae (Fig. 6.1). No obvious phylogenetic or ecological features link these families, however, so the significance of this pattern (if indeed it holds up to future examination) remains ambiguous.

Simple explanations for the distribution of variation among beetles appear to be lacking. For example, within the tenebrionids and chrysomelids there appears to be a positive relationship between C-value and chromosome number, but there are also notable (and common) cases of substantial variation in DNA content despite karyotypic constancy (Juan and Petitpierre 1991; Petitpierre et al. 1993). Simple changes in the karyotype therefore seem unlikely to explain genome size variation in these families or

4

A genome size of 3.7pg was reported for the leaf beetle *Chrysolina carnifex* (Family Chrysomelidae) by Petitpierre et al. (1993), but this seems likely to reflect a case of unrecognized polyploidy or measurement error (e.g., diploid spermatid precursors measured instead of haploid spermatids), given that none of the other 70 species of chrysomelids measured in any study (including the present one) exceeds 2pg. The mean for the entire family is only 0.78pg \pm 0.05, and all six of *C. carnifex*'s congeners studied so far have genome sizes between 0.5 and 1.2pg. For these reasons, and pending independent confirmation with measurements on cells of known ploidy (e.g., diploid haemocytes or haploid spermatozoa), this extreme outlier has been omitted from the present analyses.

among beetles at large. Phylogenetic patterns of genome size change are also difficult to discern in this order, since the relationships among families remain poorly resolved. It does bear mentioning that both of the two largest suborders are represented in the current dataset – the Adephaga by the families Carabidae, Dytiscidae, and Gyrinidae, and the larger Polyphaga by the remaining families – but that no obvious distinctions exist between them. Of the three aquatic families represented here, the Dytiscidae has a large mean C-value, the Hydrophilidae an intermediate one, and the Gyrinidae one below the beetle average (Fig. 6.1). As such, there does not appear to be any linkage between aquatic lifestyle and large or small genome size. It may be significant that the family with the largest average genome size, the Staphylinidae, is flightless, but pending the availability of further data a proper comparative test of this association cannot be undertaken.

In their discussion of genome size variation among members of the Chrysomelidae, Petitpierre and Juan (1994) note that “all species of *Timarcha* and *Chrysolina* possess genome sizes higher than 0.6pg and most of them have only one generation per year whereas *Leptinotarsa decimlineata*, *Gastrophysa polygoni*, *Chrysomela populi* and *Prasocuris junci* whose genomes sizes are lower than 0.5pg have more than one generation per year”. They then ask, “Are the smaller genomes of this second group somehow associated with the existence of several generations per year?”. The link between developmental rate and genome size as expressed in plants and amphibians was discussed in some detail in Chapter 4. In this case, Petitpierre and Juan’s (1994) query (which remains unanswered for want of more data) is similar in principle to the situation found in ephemeral or annual plants, whose genomes must be small in order

for development to be completed in a single season. In the early stages of the present study, it was hypothesized that rapidly-developing pests or invading species may be found to possess smaller genome sizes than other beetles (an analogy with weeds in this case). However, this did not appear to be the case, since highly successful imported lady beetles like *Coccinella septempunctata* and *Propylea quatuordecimpunctata* do not have noticeably smaller C-values than their endemic North American counterparts (Appendix 12). The pest Japanese beetle *Popilla japonica* also does not have particularly small genome size, suggesting that a below-average C-value is not a prerequisite for successful invasion.

Intraspecific variation in genome size has been reported for some tenebrionid beetles, most notably in *Tribolium*, but this was relatively minor and contributed far less to the total variance in the measurements than error within individual samples (Alvarez-Fuster et al. 1991). Within one species, *T. anaphe*, the variation reached 25% (Alvarez-Fuster et al. 1991), but this seems likely due to experimental error since these were all animals from the same laboratory strain, and not from geographically variable populations as in the more convincing examples of intraspecific variation (moreover, the *absolute* variation amounted to a mere 0.07pg!). In a more detailed study, Palmer and Petitpierre (1996) examined individuals of the tenebrionid *Phylan semicostatus* from eight island populations in the Mediterranean. In addition to genome size, they provided a set of morphometric measurements as an indicator of body size. They found statistically significant differences in both genome size and body size among populations, and a negative correlation between these two parameters. To explain this counter-intuitive relationship, Palmer and Petitpierre (1996) invoked a slowing of development by

larger DNA contents which ultimately results in smaller adult sizes. This interpretation is highly suspect, however, given that the total range in genome sizes was from 0.226 to 0.268pg. That a maximum difference of only 0.04pg of DNA could exert a significant impact on growth rate and body size is extremely improbable, to say the least.

Overall, there appears to be no link between DNA content and body size in beetles. Taking the mid-range body sizes for all the families included in the present dataset (from White 1983), no correlation is found with body size ($p > 0.19$). In a more taxonomically focused study, no correlation with body size was found among a small number of species in the genus *Tribolium* (Alvarez-Fuster et al. 1991), nor in tenebrionids (Juan and Petitpierre 1991) or chrysomelids (Petitpierre and Juan 1994) taken as a whole. Similar trends are also already apparent in some of the less well-studied families. For example, the diving beetle *Dytiscus dauricus*, the largest beetle species in the present dataset, has the smallest C-value among dytiscids and one roughly equal to the overall beetle average. The genome of the small whirlygig beetle (*Gyrinus* sp.) contains more than twice as much DNA as that of the large gyrid (*Dineutus* sp.). The crypt (*Cryptolaemus montrouzieri*), one of the smallest coccinellid examined here, has the largest genome in the family (Appendix 12), and the overall relationship in lady beetles is decidedly non-significant using available body size data for 15 species ($p > 0.94$).

This lack of association with body size does not mean that cell size is unrelated to genome size in beetles, however. In the best example available to date, positive correlations have been found between genome size and spermatid size (dry area) within and among genera of both tenebrionids (Juan and Petitpierre 1989, 1991; Alvarez-Fuster et al. 1991) and chrysomelids (Petitpierre et al. 1998).

In general, it is evident that beetle genome sizes are relatively constrained, with only a small number of species known to exceed a C-value of 1pg. No obvious explanations for the existing variation are forthcoming, however, and a great deal of additional data must be collected before the patterns among species and families can be deduced conclusively. The ease of collection and sheer diversity of beetles makes them a prime target of such an analysis, and it is rather surprising that very few authors have endeavoured to undertake such a project.

Collembola

As it stands, almost nothing is known of genome sizes in this primitive group of hexapods, and the estimate of 0.24pg for *Sinella caeca* performed as part of the present study is the only collembolan C-value so far available. Whether this small genome size is typical of collembolans and perhaps indicative of a small ancestral hexapod C-value remains to be determined, but these are clearly interesting possibilities worthy of continued investigation.

Dermaptera

The C-value of the European earwig (*Forficula auricularia*), estimated in the present study at about 1.4pg, is the only genome size available for a dermapteran. Like some of the beetles described above, this is an imported species that has achieved great success in North America, but which does not have a small genome.

Diptera

Aside from beetles, dipterans are the best-studied group of insects in terms of genome size variation, with roughly 65 estimates covering 12 families available for this order (Appendix 11). However, 41 of these are from the Family Culicidae (mosquitoes)

and another 12 are from the Family Drosophilidae (fruit flies). This is perhaps not surprising, given the importance of drosophilids as molecular models and of mosquitoes as conveyers of the pathogens causing malaria and dengue and yellow fevers. However, this means that the taxonomic coverage of this group is quite limited, and that insights from this order are largely restricted to patterns found within these two families (unlike with the beetles, no effort was made to correct this in the present study). Fortunately, these are well-studied groups with regards to basic biological features, such that relevant comparisons are possible despite the narrow taxonomic focus of the current dipteran C-value dataset.

Dipteran C-value estimates range approximately 20-fold, from 0.09pg in the hessian fly, *Mayetiola destructor*, to 1.9pg in the mosquito *Aedes zoosophus*. The mean genome size for all flies studied to date is $0.71\text{pg} \pm 0.05$. The distribution among the families studied thus far is depicted in Figure 6.2, but again most of the data currently available are from only two of these families.

Genome sizes in the Family Drosophilidae average only about $0.25\text{pg} \pm 0.02$, and range from 0.15pg in *D. simulans* to 0.4pg in *D. nasutoides* (Appendix 11). The genome sizes of mosquitoes (Family Culicidae) are generally larger and more variable than those of drosophilids, ranging from about 0.25pg in several species of *Anopheles* to 1.9pg in *Aedes zoosophus*. Thus, mosquito genomes display almost the entire range in C-values found in dipterans, although at $0.94\text{pg} \pm 0.06$ the average is higher than for most groups of flies. Moreover, the Culicidae is the only dipteran family currently known to contain members with genome sizes larger than 0.75pg (roughly 3/4 of mosquitoes exceed this level, in fact). There is no correlation between genome size and chromosome number in

mosquitoes, given that all species examined so far have a diploid number of $2n = 6$, although chromosome size is positively related to genome size in this group (Rao and Rai 1990). In this family, more primitive forms possess smaller DNA contents (Rao and Rai 1990; Rai and Black 1999), suggesting (but of course not proving) a trend towards DNA increase during the evolution of mosquitoes.

In addition to their relatively large interspecific range (for dipterans), mosquitoes have been singled out as displaying substantial intraspecific variation in some cases. In the first example, Rao and Rai (1987) reported more than 30% variation within *Aedes albopictus*, with island populations in Polynesia having smaller genome sizes than those collected on the Asian mainland (Rao and Rai 1987). These differences in DNA content among populations were also associated with variation in chromosome size (Rao and Rai 1987), and a subsequent reassociation kinetic analysis showed such differences among *A. albopictus* strains to correspond to variation in amounts of highly repetitive DNA (Black and Rai 1988).

In another interesting corroboration of this reported variation within *A. albopictus*, Ferrari and Rai (1989) compared the genome sizes, development times, and wing lengths (a proxy for body size) in 10 different strains. The removal of a single outlier strain from the analysis revealed significant positive correlations between genome size and both development time and wing size (Ferrari and Rai 1989)⁵. In addition, both development

5

The outlier in this case, the “Calcutta” strain, had the smallest genome but a relatively long development time (Ferrari and Rai 1989). As discussed in Chapter 4, it is not uncommon for animals with low DNA contents to show a range of developmental rates, whereas large genomes are found exclusively in slow developers. The relationship with wing size in this case is much more convincing in the case of mosquitoes than in the beetle example described previously, for two reasons. First, the variation in this case is much greater and therefore more likely to exert an observable effect on morphology. Second, differences in wing size among populations of

time and C-value were positively associated with the number of years that the strains had been kept in culture, suggesting some positive selection for slower growth under laboratory conditions that also affected genome size, or (more likely, in my opinion) a release of genome size/developmental rate constraints active under natural conditions but not in the lab (Ferrari and Rai 1989). In any case, in this study it appeared that “most of the phenotypic variation in genome size and development time was generated in the laboratory” (Ferrari and Rai 1989).

The example of laboratory-reared *A. albopictus* not only illustrates the link between genome size and developmental and morphological traits, but also hints at the existence of stabilizing constraints on genome size resulting from these associations that can only be escaped when ecological pressures change. In keeping with this, Kumar and Rai (1990) later examined a total of 37 populations of *A. albopictus* sampled from throughout its global range, and found increases in DNA content in newly colonized areas. However, no clear geographic clines were apparent, and they proposed instead that “the variation in DNA content among populations of *Ae. albopictus* ... is due mainly to repetitive DNA sequences that are under rapid change in response to the microenvironment” (Kumar and Rai 1990). If the repetitive DNA sequences in question are transposable elements activated and/or released from constraints by changes in the environment (as they are in *Drosophila*; Vieira et al. 2002), then this would provide a prime example of a causative hierarchical interaction between the subgenomic,

conspecific flies is known to be affected by variation in cell size (e.g., Gilchrist and Partridge 1999), whereas generalized body size in beetles may be determined primarily by differences in cell number (hence the lack of relationship with genome size across beetle species with clearly divergent C-values; see above).

organismal, and ecological levels. In the reverse direction along this hierarchy, the action of transposable elements has been invoked to explain the ability of mosquitoes to colonize new environments and evolve resistance to pesticides (Bensaadi-Merchermek et al. 1994).

Embiidina

Members of the Order Embiidina (= Embioptera) are commonly known as “webspinners” because of their capacity to produce silk from glands in their fore-tarsi. On superficial morphological grounds, webspinners bear similarities to both earwigs (Order Dermaptera) and stoneflies (Order Plecoptera), but these groups are not closely related. This small order of primarily tropical insects is rarely encountered, and has therefore not previously been examined from the standpoint of genome size. As part of the present study, measurements on haemocytes and sperm of the Brazilian species *Biguembia multivenosa* revealed a C-value of 1.45pg. As mentioned in Chapter 5, webspinners possess a semi-unique sperm nucleus morphology otherwise found only in the stick insects (Order Phasmida).

Ephemeroptera

Unlike webspinners, mayflies are commonly encountered in North America, but this group has also been neglected in previous genome size studies. Two species of mayflies from two different families have been analyzed in the present study, *Ephemerella subvaria* (Family Ephemerellidae) at 0.72pg and *Ameletus* sp. (Family Ameletidae) at 0.76pg.

Hemiptera

The previously available data on hemipterans consist of estimates for four

representatives of the Suborder Heteroptera and 35 from the Suborder Homoptera. In total, this covers six families, with the only homopterans being aphids. The opposite tack was taken in the present study, in which four homopterans (from three unstudied families) and 24 heteropterans (11 families) were added to the dataset. The mean for all hemipterans studied so far is $0.91\text{pg} \pm 0.12$. For heteropterans, it is $1.18\text{pg} \pm 0.18$, and for homopterans $0.72\text{pg} \pm 0.16$, although obviously this latter value and that for the overall mean are heavily biased by the predominance of aphids in the dataset. The C-values of aphids (families Adelgidae and Aphididae) range 5-fold, from 0.18pg in *Eoessigia longicauda* to 0.89pg in *Gypsoaphis oestlundii*, with a mean of $0.51\text{pg} \pm 0.03$. The four non-aphid homopterans display a mean C-value of $2.5\text{pg} \pm 1.4$, and range in genome size from about 1.1pg in the leafhopper *Athysanus argentarius* to 6.7pg in the dog day cicada *Tibicen canicularis*. At least one other cicada is believed to have a similarly large genome size ($\sim 4.5\text{pg}$; J.S. Johnston, personal communication). The range in heteropterans is roughly equal to that for homopterans, from 0.3pg in the water strider *Gerris remigis* to 5.4pg in the milkweed bug *Oncopeltus fasciatus*. The patterns of distribution among the hemipteran families studied so far are illustrated in Figure 6.3.

In general, only the aphids have obviously constrained genome sizes, with none of the 35 studied species from either family exceeding 1pg . Nearly all the other hemipteran families, despite being much less heavily surveyed, contain at least one member with a C-value higher than 1pg (and more than 60% of the non-aphid species do). The primary question to be addressed therefore relates to the reasons for these constraints among aphids, but not in other hemipterans (particularly other homopterans). At a proximate level, it is relevant that aphid genomes contain very little highly repetitive DNA (Ma et al.

1992). This may suggest active mechanisms operating to maintain small genome sizes, which could relate in ultimate terms to the need for rapid development in these parthenogenetic animals. Thus,

The ability of aphids to form enormous populations in a short period of time is a result of apomictic parthenogenesis, telescoping generations in the ovaries of the mother, and short (5-7 days) nymph to adult development time. All of these adaptations reflect selection for rapid development during aphid evolution. If there is a causal relationship between genome size and development time, selection for rapid development time may have maintained the aphid genome relatively free of highly repetitive DNA. (Ma et al. 1992).

Another possible constraint on aphid genome size arises as a result of the significant positive correlation with body size in this group (Finston et al. 1995). Interestingly, there is evidence of “quantum shifts” in C-value among aphids that is unrelated to chromosome number changes (Finston et al. 1995). Among the remaining hemipterans, as with beetles, there is no obvious relationship between genome size and body size. Although the cicada has both a large body size and a high C-value, many counterexamples are apparent. For example, the large water strider *Gerris remigis* has a smaller genome than its more diminutive congener, and indeed the smallest genome yet reported for a heteropteran. Most notably, the largest hemipteran in North America, the giant waterbug *Lethocerus americanus*, has a genome size smaller than that of some aphids. (These examples should not be taken to imply that aquatic hemipterans have particularly small genomes. As with the beetles, there is no apparent relationship between genome size and aquatic versus terrestrial lifestyle; see Fig. 6.3). It would therefore seem that non-aphids are generally free of such developmental and

morphological constraints, thereby allowing their genomes to expand in size well beyond the threshold experienced by their aphid relatives.

There is some minor evidence for intraspecific variation in aphids. In their study of *Schizaphis graminum*, Thindwa et al. (1994) found specimens raised on sorghum to have lower measured DNA contents than individuals from the same biotypes reared on wheat or johnsongrass. However, it seems plausible that this is an example of measurement error. The aphids measured in this study were simply homogenized, so it is possible that some botanical compound in the food confused staining with propidium iodide (cf. Price et al. 2000). Alternatively, perhaps certain rearing conditions promote more or less endopolyploidization of specific tissues, which would register as a difference in mean DNA content. Obviously, the mere use of flow cytometry does not rule out the role of experimental errors of this kind. Thindwa et al. (1994) propose the existence of unique growth conditions on sorghum that may select for faster development and lower DNA content, but the absolute differences in C-value were so small (on the order of 0.02 to 0.05pg) that there is unlikely to be any biological significance to this variation, even if real.

Hymenoptera

Including the seven new wasp C-value measurements made here, only 14 species of Hymenoptera have been studied to date. In total, this covers seven families, with only one representative each for ants (Family Formicidae), leaf-cutting bees (Family Megachilidae), and several groups of wasps (families Encyrtidae, Pteromalidae, and Vespidae). Two species of honey bees (Family Apidae) and seven parasitic wasps (Family Braconidae) have also been studied. As a whole, the mean for this limited

dataset is $0.25\text{pg} \pm 0.04$, with a range from 0.1 to 0.6pg. The largest genomes are found in the red fire ant, *Solenopsis invicta* (0.62pg), and the parasitic wasp, *Copidosoma floridanum* (0.57pg). The other parasitic wasps (Family Braconidae) all have genomes of 0.16pg or less (that is, all are smaller than *Drosophila melanogaster*). Thus, there may be an association between small genome size and a parasitic lifestyle, at least in this family. Both honey bee species (Family Apidae) appear to have particularly small genomes as well (0.19pg). With such a limited dataset, it is difficult to assess any broader patterns, although there are many questions that could be addressed effectively with this group as a model. These include issues of social versus solitary species, parasitic versus non-parasitic forms, and many questions relating to different behavioural and morphological characters among the widely variable ants.

Isoptera

Two species in the termite Family Rhinotermitidae were measured in the present study, representing the first members of the Isoptera to be evaluated for genome size. The estimated C-values were similar for the two species, being 1.14pg in the eastern subterranean termite (*Reticulotermes flavipes*) and 1.01pg in the Formosan subterranean termite (*Coptotermes formosanus*). These are noticeably larger than the C-values found in hymenopterans, suggesting that sociality *per se* does not require an especially small genome as a result of requirements for rapid development or other related features.

Lepidoptera

Moths and butterflies have been collected and studied with great fervour for centuries and throughout the world, but this has not included measurements of nuclear DNA content. Before the present study, only five lepidopterans (all moths) had been

evaluated in this manner. This included four commercially important species, the useful silkworm moth *Bombyx mori* at 0.5pg and Chinese oak silkmoth *Antheraea pernyi* at 1pg, and the pest gypsy moth *Lymantria dispar* at 1pg and wax moth *Galleria melonella* at 0.5pg. The geometrid *Lycia pomonaria* provides the last and largest C-value in this dataset at about 1.5pg.

The present study included estimates for two butterflies (2 families) and 50 moths (8 families), most of them collected by ultraviolet light near Guelph, ON (by Paul Hebert) or by hand net at Turkey Point, ON (by TRG). With these new data, the lepidopteran dataset now covers 14 families, with C-values ranging from 0.29pg in the monarch butterfly *Danaus plexippus* to 1.9pg in the geometrid moth *Euchlaena irraria*. The only other butterfly studied, the African satyrid *Bicyclus anynana*, has a C-value of 0.49pg, suggesting that butterflies may have smaller genomes on average than moths (Fig. 6.4), although this is obviously little more than speculation in light of the current dataset. The average for all moths is $0.68\text{pg} \pm 0.04$.

No distinct patterns are obvious on the basis of the current data. For example, it is notable that the largest moth so far studied (by far), the modest sphinx *Pachysphinx modesta*, has a below-average C-value of only 0.46pg. Other relatively large sphingids (though smaller than *P. modesta*) have genome sizes between 0.58 and 0.68pg, and the very small plume moths (Family Pterophoridae) have C-values between 0.6 and 0.9pg. These examples may be taken to suggest that body size is not correlated with genome size in moths, but note that no micromoths were included in the present survey. Other potential life history traits (e.g., diurnal versus nocturnal lifestyle, relative activity levels, etc.) may be of interest from the perspective of genome size, but it does bear mentioning

that no family of moths here analyzed possessed unilaterally small or large genome sizes, and indeed most seemed quite variable (Appendix 12). In this regard, it may be difficult to discern any patterns within or among families without the collection of large amounts of data from a diverse sample of species.

Mantodea

Several species of mantids in the genus were measured by Hughes-Schrader (1951, 1953), but her results were given only in arbitrary and could not be converted to absolute genome sizes because no internal standard was included. As such, the C-value for the praying mantis (*Mantis religiosa*), estimated in the present study at 3.15pg, is the only datum currently available for this order.

Odonata

The dragonfly *Aeshna confusa* and the damselfly *Calopteryx splendens* are the only two members of this order to have been studied for genome size, and both have relatively large genomes (2.2 and 1.5pg, respectively).

Orthoptera

Four families of orthopterans had been examined for variation in genome size prior to the present study. These included the short-horned grasshopper Family Acrididae (37 species), the monkey grasshopper Family Eumasticidae (1 species), the cricket Family Gryllidae (4 species), and the camel cricket Family Gryllacrididae (1 species). Four families were included in the present study, including four species of grasshoppers (Family Acrididae), four crickets (Family Gryllidae), two camel crickets (Family Gryllacrididae) and five katydids (Family Tettigoniidae). Only two species were repeated from prior measurements, and the katydids included in the present analysis are the first

members of this family to be examined.

With all 56 orthopteran species considered together, the mean for the order is $8.2\text{pg} \pm 0.5$, with a roughly 11-fold range (1.55 to 17pg). However, this calculation is strongly biased towards acridid grasshoppers, which make up nearly 3/4 of the current dataset. Moreover, there are substantial differences in genome size distributions among families, as shown in Figure 6.5. Thus, the Acrididae average about $9.5\text{pg} \pm 0.5$, with a range more limited to higher C-values (4.5 to 17pg). Indeed, the highest C-value in an orthopteran, which is found in the mountain grasshopper *Podisma pedestris*, is also the highest for any insect. At the opposite extreme are the Gryllidae, with a mean C-value of only $1.9\text{pg} \pm 0.1$, ranging only from 1.5 to 2.2pg. The single representative of the Family Eumasticidae, the grasshopper *Warramaba virgo*, has a genome intermediate between these two at 3.9pg. Camel crickets (Family Gryllacrididae) are much more similar to the Acrididae than to the Gryllidae, displaying genome sizes from 5.2 to 9.6pg and a mean of $7.3\text{pg} \pm 1.3$. Finally, the katydids (Family Tettigoniidae) display a wide range in C-values, from 2.5 to 10.1pg, and a relatively high mean genome size ($6.2\text{pg} \pm 1.3$).

No clear phylogenetic patterns obtain in this dataset. Thus, the Suborders Caelifera (Acrididae and Eumasticidae) and Ensifera (Gryllacrididae, Gryllidae, and Tettigoniidae) both contain members with relatively low and very high C-values. The Gryllidae, and perhaps also the Eumasticidae, have the smallest C-values but their closest relatives have some of the largest. It is possible that data from additional families will clarify some currently unrecognizable patterns, but it may also be the case that the specific biological features of each group act to shape genome size independent of phylogenetic associations among families. It does bear mentioning that the Acrididae,

which have the largest genomes of any insect group studied to date, seem particularly tolerant of modulations in DNA content, as evidenced by the frequent occurrence of B chromosomes in this group. These are not without biological effects, however, and the presence of B chromosomes has a demonstrable impact on developmental rate (Harvey and Hewitt 1979) and gamete production (Suja et al. 1986; Hewitt et al. 1987) in grasshoppers. In some cases, supernumerary elements or chromosome fusions contribute to significant differences among geographic races of the same species (e.g., Rees et al. 1978; Westerman et al. 1987). However, it is also the case that most variation in haploid genome size among acridids comes without substantial differences in chromosome numbers (e.g., John and Hewitt 1966; Rees et al. 1978), so polyploidy is not the main determinant of DNA content in this group. Instead, variation in heterochromatin content may explain these size of these weightiest of insect genomes (Belda et al. 1991).

Phasmida

Stick insects of the Order Phasmida (= Phasmatodea) are intriguing for a variety of reasons. Most obviously, their uncanny resemblance to the vegetation upon which they live as adults is among the most impressive examples of mimicry in the animal kingdom. In some species (e.g., *Extatosoma tiaratum*), the eggs resemble plant seeds that ants carry into their nests (where they are afforded protection from predators), and which then hatch into nymphs that are mimics of their hymenopteran hosts. The Phasmida also contains the largest (or at least the longest, at over 36cm) insect in the world, *Pharnacia kirbyi*. Many species undergo parthenogenetic reproduction, which makes them interesting from the standpoint of reproductive and developmental biology.

Phasmids are also of interest from a cytogenetic point of view. For example,

some species of stick insects exhibit interspecific hybridization and polyploidy (e.g., Marescalchi et al. 1990; Marescalchi and Scali 2001). In addition, the leaf insect *E. tiaratum* (Family Phasmatidae) may have one of the largest genome sizes yet reported for an insect ($1C = 8.0\text{pg}$)⁶. Whether large genomes such as this are typical of phasmids remains to be seen. Before the present study, the genome sizes of representatives from only one genus of stick insects (*Bacillus* spp., Family Phasmatidae) had been measured, ranging about $1C = 2$ to 3pg in size (four species and seven subspecies in total; Appendix 11).

The northern walkingstick (*Diapheromera femorata*, Family Heteronemiidae) is the only species of stick insect found in Canada, where it is restricted primarily to southern regions of Ontario and Québec (Vickery and Kevan 1985). Adults of *D. femorata* are defoliators of oak trees, and are therefore considered potential, and occasionally actual, ecological pests (Wilson 1971; Giese and Knauer 1977; Vickery and Kevan 1985). The present study provided an estimate of the genome size of *D. femorata*, thereby marking the first such examination of a North American stick insect, and only the third genus (and second family) so far studied from the order Phasmida. The genome size of *D. femorata* was estimated at 2.55pg , which is within the range of values for representatives of the genus *Bacillus*. The possibility that phasmid genome sizes are consistently large across different families is supported by this observation, though much

6

However, note that this value for *Extatosoma tiaratum* was originally reported to be $1C = 24.0\text{pg}$ by Nagl and Schäffner (1981) but had to be corrected because of an inaccurate C-value for the standard, *Acheta domestica*. Based on the fact that neural cells were measured, there may also be reason to believe that 8.0pg actually represents the $2C$ value. In some cases, neurons can be highly polyploid, as in the remarkable example of giant neurons (up to 1mm in diameter!) in the mollusc *Aplysia californica*, which are about 200,000-ploid (Lasek and Dower 1971).

work remains to be done before any generalizations can be made regarding a “typical” genome size for the order as a whole.

European populations of *Bacillus atticus* consist of both diploid and triploid parthenogenetic strains, among the former of which there is significant latitudinal variation (~20%) in nuclear DNA content, decreasing gradually from east to west (Marescalchi et al. 1998). By contrast, *D. femorata* is bisexual, and the consistent DNA contents of male and female somatic cells and sperm indicate a likely diploid state in this species. Geographical variation within *D. femorata* was not evaluated in the present study, but at least within the single population sampled here there was no detectable intraspecific variation in genome size. In both *B. grandii* and *B. rossius*, the only two bisexual species of *Bacillus* studied, the sex determination system results in females (XX) with one more chromosome and about 10-20% more DNA than males (XO) (Marescalchi et al. 1990). Again in contrast to these European species, *D. femorata* showed no differences in haemocyte DNA content between males and females (*t*-test, $p > 0.30$).

Plecoptera

The four species of stoneflies included in the present study are the first of this order to be examined for genome size. In this case, stonefly genomes ranged from 0.36 to 2.15pg, the smallest being found in a winter stonefly of the Family Capniidae and the largest in *Acroneuria abnormis* of the Family Perlidae (Appendix 12).

Siphonaptera

Only the cat flea, *Ctenocephalides felis*, has been examined for genome size in the Order Siphonaptera. Its C-value is here estimated to be 0.27pg, which is at the lower end of haploid DNA contents for insects. Whether this small genome size is typical of fleas

and perhaps related to a parasitic lifestyle cannot be determined without a broader sampling of this order. It may be relevant in this context that the cat flea has a notorious capability for rapid reproduction (and therefore infestation).

Thysanura

Silverfishes and firebrats (more generally, “bristletails”) are members of the wingless Order Thysanura, the only group in the Subclass Dicondylia included in the present genome size dataset (all the other orders, with the exception of the non-insect Collembola, are in the Subclass Pterygota). The genome size of the firebrat, *Thermobia domestica*, had previously been estimated to be about 3.1pg, and the present study added data from two more thysanurans, the silverfishes *Ctenolepisma quadriseriata* and *Lepisma saccharina*, both with C-values around 5.2pg. All three species are members of the Family Lepismatidae, such that the taxonomic coverage of this group remains fairly narrow. Nevertheless, it is very interesting that this primitive order of insects has such large C-values, in contrast to the apparent situation in the Collembola. Again, inferences of ancestral states based on comparisons of extant taxa can be fraught with difficulty, as seen in the fishes, amphibians, and now also the insects.

Trichoptera

The single caddisfly studied thus far, *Pycnopsyche* sp., was estimated in the present study to have a C-value of 0.66pg.

Patterns of variation in arachnids

General patterns in spiders

The spider C-values included in the present dataset range nearly eight-fold, from 0.74pg in the long-jawed orb-weaver *Tetragnatha elongata* to 5.7pg in jumping spiders of

the genus *Habronattus* (Appendix 13). The mean for all spiders surveyed was $2.4\text{pg} \pm 0.1$. Even with the narrow taxonomic coverage of the present dataset, it is apparent that some families are more constrained in their genome size variation than others, as shown in Figure 6.6. In some cases, the actual extent of variation cannot be assessed due to low number of species analyzed, but in others such as the Family Araneidae, which is represented here by 17 species, the consistency of C-values is considerable (Fig. 6.6).

Based on the current dataset, which is large in terms of absolute species numbers but still very limited in overall taxonomic representation, no obvious associations can be found with any of the characteristics that define the differences among spider families. Thus, no clear patterns can yet be discerned in relation to such features as “web-building/structure, general body size, eye arrangement, type of prey, etc.” (D.P. Shorthouse, personal communication), nor are there any obvious phylogenetic patterns based on the cladistic hypothesis of Coddington and Levi (1991).

In the current dataset, families such as the Anyphaenidae (ghost spiders), Clubionidae (sac spiders), Dyctinidae (meshweavers), Pholcidae (cellar spiders), and Tetragnathidae (long-jawed orb-weavers) appear to have small genomes, while the Agelinae (funnel weavers), Amaurobiidae (hackledmesh weavers), Dysderidae, Lycosidae (wolf spiders), Salticidae (jumping spiders), and Thomisidae (crab spiders) have the largest C-values. Other families – such as the Araneidae (orb-weavers) and Theridiidae (cobweb weavers), which are well-represented in this dataset – appear intermediate in terms of genome size. Again, whether there is an association between genome size and predation strategy (e.g., the construction of elaborate webs as in the Araneidae and Theridiidae, a webless lie-in-wait strategy as in the Thomisidae, or active

hunting as in the Lycosidae and Salticidae) remains to be seen. With the addition of more data, some interesting patterns are sure to emerge. Data on other groups of arachnids (e.g., “tarantulas”, scorpions, harvestmen, mites, and ticks) would also be very useful from a comparative standpoint.

A note on the spider mite

Both mites measured here possessed small genomes – the water mite 0.19pg and the two-spotted spider mite a minuscule 0.08pg. This latter value is one of the smallest known for an arthropod, being only about 40% of that of *Drosophila melanogaster*⁷. Interestingly, this species is endowed with a phenomenally rapid developmental rate, passing from egg to egg-laying adult in about one week under warm conditions, despite passing through larval, protonymph, and deutonymph stages (Osborne et al. 1999). That the brevity of the spider mite’s developmental program is linked to its tiny genome size seems very plausible.

Patterns of variation in myriapods

In total, the genomes sizes in this group ranged from 0.28 to 2.14pg, with a mean of $0.78\text{pg} \pm 0.16$. Both small and relatively large C-values were found in centipedes (Class Chilopoda, 0.39 to 2.14pg) and millipedes (Class Diplopoda, 0.28 to 1.4pg). The two most well known centipedes, the common house centipede (*Scutigera coleoptrata*)

7

Miodrag Grbic (University of Western Ontario) and I have attempted to present the suggestion (which so far remains unpublished!) that this animal be targeted for a major genome sequencing project. Again, its genome size is less than half that of *Drosophila melanogaster*, and is even smaller than that of *Caenorhabditis elegans*, which means that a genome sequencing and mapping project for this species would be among the simplest of any yet attempted for a multicellular organism. Given the economic significance and long list of intriguing biological features of this species, there is little doubt that the detailed study of this tiny animal will pay large dividends.

and the rock

centipede (*Lithobius forficatus*) had the largest genomes in the sample (2.03pg and 2.14pg, respectively). With the exception of one unidentified millipede at 1.4pg, all other myriapods analyzed here had C-values less than 0.9pg (Appendix 20).

Developmental complexity in insects: more on metamorphosis

Aside from a few cases within families, body size appears largely unrelated to genome size in insects. No divisions between aquatic and terrestrial lifestyles are obvious either. In fact, the patterns of variation among insect orders have not been explained by any of the relationships heretofore considered in this (or any other) discussion of insect C-values. However, with the exception of a few small examples, one notable feature dealt with in previous chapters has not yet been addressed here, namely *development*.

As seen in Chapter 4, developmental parameters may be of substantial importance in the context of genome size evolution in some groups. Developmental rate, the simpler side of the ontogenetic coin, has been found to correlate with genome size within mosquitoes, for example, and the role of DNA content modulation in development is made obvious by the examples of endopolyploidy discussed above. Yet, these seem unlikely to explain the variation in DNA content among major groups of insects, and it may be the alternate approach of developmental *complexity* that is relevant in insects, as it is in amphibians (Chapter 4).

The various insect orders are divided into three broad categories according to the nature of their developmental programs (particularly with reference to wing development). The Apterygota display *ametabolous* development, such that they emerge as miniature adults and undergo no metamorphosis whatsoever. The Exopterygota

possess a *hemimetabolous* developmental program, in which they grow via a series of nymphal moults and display only partial metamorphosis. Finally, the Endopterygota, which display *holometabolous* development, have distinct larval stages and proceed through a complete metamorphosis. In light of these divisions, the insects are similar to the amphibians (which may be biphasic, direct-developing, or neotenic), although insect metamorphosis, when it occurs, is considerably more intensive than even that of frogs.

Before the addition of data provided in the present study, the basic pattern observed in insect genome size variation was as follows. Most of the known insect genome sizes fell within the range of 0.2 to 1.5pg, with the notable exceptions including members of the Orthoptera (1.5 to 17pg), Odonata (1.5 to 2pg), Hemiptera (0.2 to 6.2pg), Phasmida (2 to 8pg), and Blattaria (2 to 3pg). The remaining orders (Diptera, Coleoptera, Hymenoptera, Lepidoptera), on the other hand, all appeared to consist of members with small genomes. As I pointed out before beginning the measurements in this study (Gregory 2002b⁸), the common unifying feature of the first set of large-genomed orders is that they all display hemimetabolous development lacking complete metamorphosis. The members of the holometabolous orders analyzed to that point, by contrast, appeared to

8

This paper was written in the spring of 2001, but was part of a special issue on genome size evolution that did not appear in print until more than a year later (May 2002 issue of *Genetica*, actually not made available until July). Interestingly, while preparing to write this chapter I came across a brief discussion in Barlow (1978) who cites Bier and Müller (1969) as arguing that insect orders with panoistic ovaries have genome sizes larger than 2pg, whereas those with meroistic ovaries have genomes smaller than 2pg. Since ovary types are divided largely along the same lines as metamorphic lifestyles, this could have represented a case of coincidental consilience (i.e., me being scooped). However, a closer examination of Bier and Müller's (1969) paper reveals that although they do indeed discuss differences in genome size between insects with different types of ovaries, but they do not explicitly point out a 2pg threshold. This value actually comes from the limited dataset they included (two orthopterans, two dipterans), which happened to be divided at a level between 1.0 and 2.3pg.

have genomes no larger (and almost always much smaller) than 2pg. Based on this distinction, I predicted that as-yet-unstudied hemimetabolous orders would contain members with high maximum genome sizes, while holometabolous orders would not exceed a hypothetical 2pg threshold. In part, the 230 new measurements conducted as part of the present study were done in order to test this hypothesis.

Using the much broader dataset of insect genome sizes now available, it is apparent that the 2pg threshold has held up well to preliminary empirical testing. As seen in Figures 6.7 and 6.8, insect C-values do tend to assort themselves along these developmental lines, with large genomes found exclusively in ametabolous and hemimetabolous orders. Moreover, the present study's inclusion of numerous families of holometabolous groups like beetles and moths suggests that this pattern is not the product of fortuitous sampling from small subsets of the various orders. Indeed, almost no exceptions have yet been found to this basic pattern. The only (potential) example from the present dataset is found in a leaf beetle with a reported genome size over 3pg, although this seems quite suspect, as discussed above. It is worth noting in this context that none of the other 221 species of beetles studied so far breach the 2pg threshold by more than 10% (and only two do so at all).

As with the amphibians, those groups which must develop and metamorphose quickly have some of the smallest genomes among insects (and indeed, among animals). Notable examples include the parasitic braconid wasps, and perhaps also fleas. In a more general sense, small genome size seems linked to rapid diversification, as shown by the fact that four of the five largest insect orders are holometabolous with small C-values (Coleoptera, Hymenoptera, Lepidoptera, and Diptera). Conversely, insects without

metamorphosis may possess very large genomes, and it bears noting that both spiders and myriapods, which do not undergo metamorphosis, also display high maximum C-values.

Of course, it should be emphasized that metamorphosis is not the only relevant parameter in discussions of insect genome size evolution. Aphids, as mentioned above, have incomplete metamorphosis but tiny genomes. In these and other insects, additional selective pressures such as that for small body size and/or rapid overall life cycle may contribute significantly to the resulting C-value (Ma et al. 1992; Finston et al. 1995). In keeping with this, the single collembolan so far studied, which has ametabolous – but rapid – development similarly has a small genome, whereas other ametabolous insects that do not have such rapid developmental rates (Order Thysanura) display rather large C-values. Yet again, the notion here is one of thresholds beyond which metamorphosis would be impeded, but below (or in the absence) of which genome size may be quite malleable in response to various evolutionary forces. This situation parallels that found in amphibians and plants (Chapter 4).

Although the present study has greatly improved the coverage of insect groups, many orders remain entirely unknown in terms of genome size. As such, additional opportunities to test the 2pg metamorphic threshold exist. Ametabolous orders like the Protura and Diplura may have large genomes, although other factors may warrant consideration (developmental rate, body size), as with the collembolans. Additional confirmations (or refutations, such as the case may be) of the threshold can be pursued by examining members of holometabolous orders which vary widely in biological features other than the existence of complete metamorphosis (as the body size and aquatic-terrestrial comparisons have done here). Holometabolous orders not covered so far

include scorpionflies (Mecoptera), lacewings (Neuroptera), and endoparasitic stylops (Strepsiptera). From the opposite perspective, new data from previously overlooked hemimetabolous orders can be used to test the hypothesis of high relative C-values in these groups, at least in the absence of other constraints. Prime candidates in this case include the Zoraptera, booklice (Psocoptera), biting lice (Mallophaga) and sucking lice (Siphunculata)⁹. Thrips (Order Thysanoptera) may be of particular interest in this regard, because although they are exopterygotes, they have evolved a sort of “holometabolous” development with a distinct pupal stage.

Whether or not the 2pg developmental threshold survives continued testing remains to be seen. But regardless of its fate as an explanation for the observed distribution of insect C-values, if it helps to stimulate further study of genome size variation in this maximally speciose but largely overlooked group, it will have served an important purpose.

Concluding remarks

Despite their general neglect in previous genome size surveys, it is obvious that insects and other terrestrial arthropods are of substantial interest in studies of the C-value enigma. In particular, the results of the present study have revealed a possible linkage between metamorphic lifestyle and small genome size, thereby providing an independent (though preliminary) test of the developmental complexity hypothesis proposed in Chapter 4. However, even with the relatively large sample size and variable data now available, patterns found in other groups of animals remain difficult to establish in

⁹

Sometimes the Mallophaga and Siphunculata are combined into the single Order Anoplura.

terrestrial arthropods. Most notably, there does not appear to be any obvious relationship between genome size and body size within either insects or spiders (although a weak trend may be identified among orders in considering the extremes of dipterans versus orthopterans). It may be that additional data will reveal such patterns, but it is also possible that arthropod genome size variation will require explanations not yet developed for any animal groups.

It is similarly unclear at this point what the exact phylogenetic trajectory has been for insect genome size evolution. Primitive hexapods like collembolans and thysanurans differ greatly in C-value, and illustrate the problematic nature of making simple inferences of ancestry based entirely on extant taxa. It seems plausible that there was a secondary reduction in the relevant orders when complete metamorphosis evolved, but it remains possible that the lineages that could and did become holometabolous are those that maintained an initially small C-value. From the opposite perspective, it could be that the members of hemimetabolous orders were unable to evolve complete metamorphosis because of their large genomes, or that their genomes expanded secondarily as a result of a lack of developmental constraints. Again, these same issues are relevant to genome evolution in amphibians (Chapter 4), and may also be applicable to spiders and other arthropods.

For the most part, the study of genome size evolution in insects and other arthropods (and therefore of animals at large) is only beginning. The present analysis has provided some descriptive insights into potential patterns, but the current dataset remains far too limited to resolve any of the most pressing questions in the C-value enigma. Perhaps more importantly at this stage, this discussion presented here illustrates the utility

of arthropods as subjects of genome size study, and may serve to stimulate further investigation into this most diverse of all animal groups.

Table 6.1. Summary of the various hexapod orders examined for genome size to date. The number of species for which C-values have been estimated is given for all previous studies combined and for the present study.

Order	Number of species studied previously	Number of species in the present study
Blattaria	4	3
Coleoptera	118	105
Collembola ¹	0	1
Dermaptera	0	1
Diptera	66	2
Embiidina	0	1
Ephemeroptera	0	2
Hemiptera ²		
Heteroptera	4 ³	25
Homoptera	35 ⁴	4
Hymenoptera	7	7
Isoptera	0	2
Lepidoptera	5	52
Mantodea	5	1
Odonata	2	0
Orthoptera	43 ⁶	15
Phasmida	8 ⁷	1
Plecoptera	0	4
Siphonaptera	0	1
Thysanura	1	2
Trichoptera	0	1
TOTAL	293	230

Table 6.1 notes:

- 1) Collembolans are considered members of the Subphylum Hexapoda, but not part of the Class Insecta.
- 2) The Hemiptera is here considered to consist of the Suborders Heteroptera and Homoptera, rather than granting these independent ordinal status.
- 3) Several additional species in the Family Pentatomidae were measured by Hughes-Schrader and Schrader (1956) and Schrader and Hughes-Schrader (1956), but were given in arbitrary units only and could not be converted to absolute genome sizes.
- 4) All previous homopteran estimates were from the aphid families Adelgidae and Aphididae, with the exception of one species each of scales in the families Coccidae and Diaspididae studied by Hughes-Schrader (1957). However, these latter estimates were given in arbitrary units only and could not be converted to absolute genome sizes.
- 5) Several species of mantids were measured by Hughes-Schrader (1951, 1953), but were given in arbitrary units only and could not be converted to absolute genome sizes.
- 6) Several additional species in the Family Acrididae were measured by Kiknadze and Vysotskaya (1969), but were given in arbitrary units only and could not be converted to absolute genome sizes.
- 7) Includes four subspecies of *Bacillus atticus*.

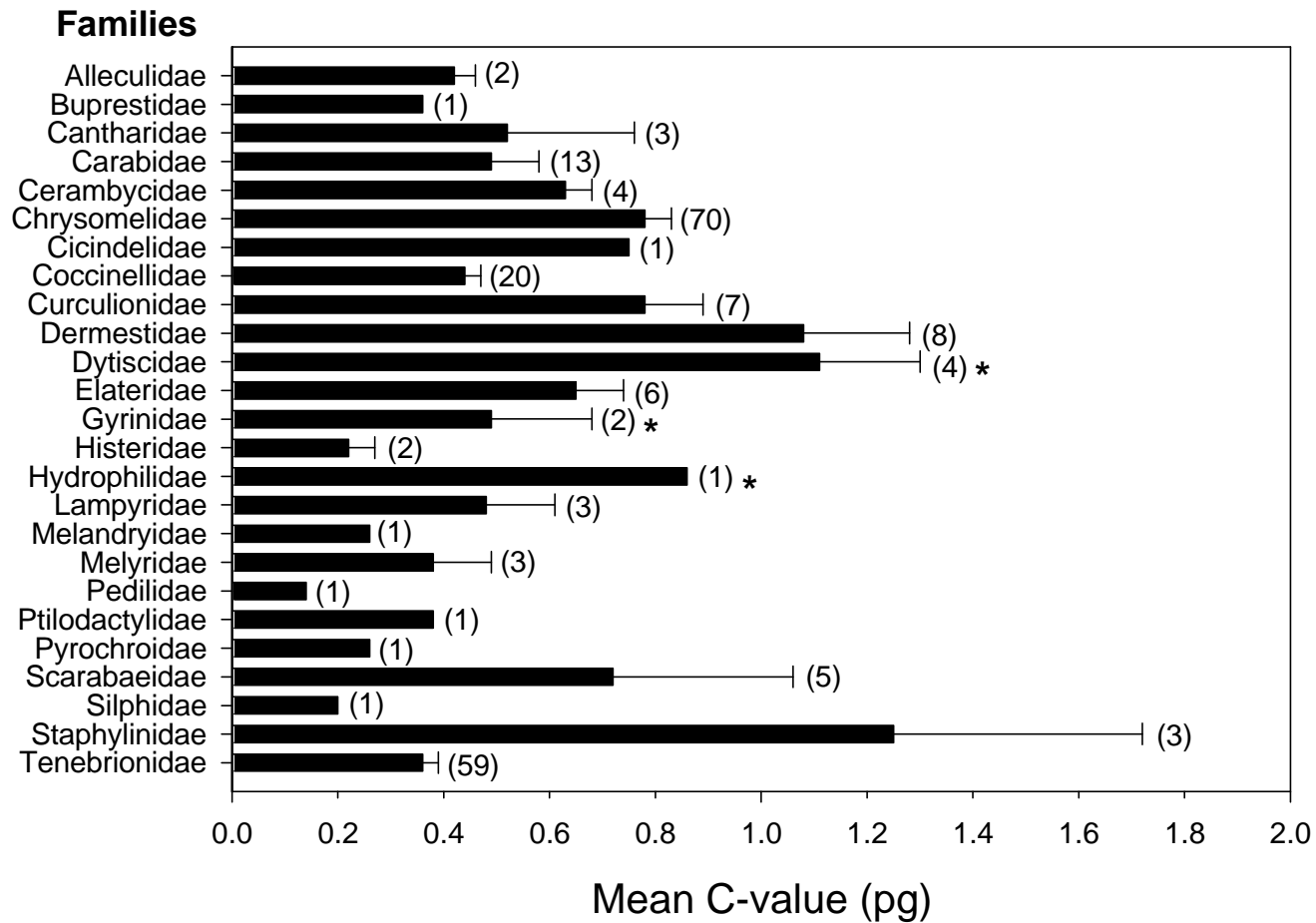


Figure 6.1. Summary of haploid genome size (C-value) variation in the 25 families of beetles (Order Coleoptera) studied to date. Families are arranged alphabetically. Numbers in brackets indicate the number of species measured per family. Error bars represent standard error. The mean C-value for the Coleoptera is $0.61\text{pg} \pm 0.03$ (range: 0.14-2.2pg), based on roughly 220 species. Aquatic families are marked with an asterisk. Genome sizes as in Appendices 11 and 12.

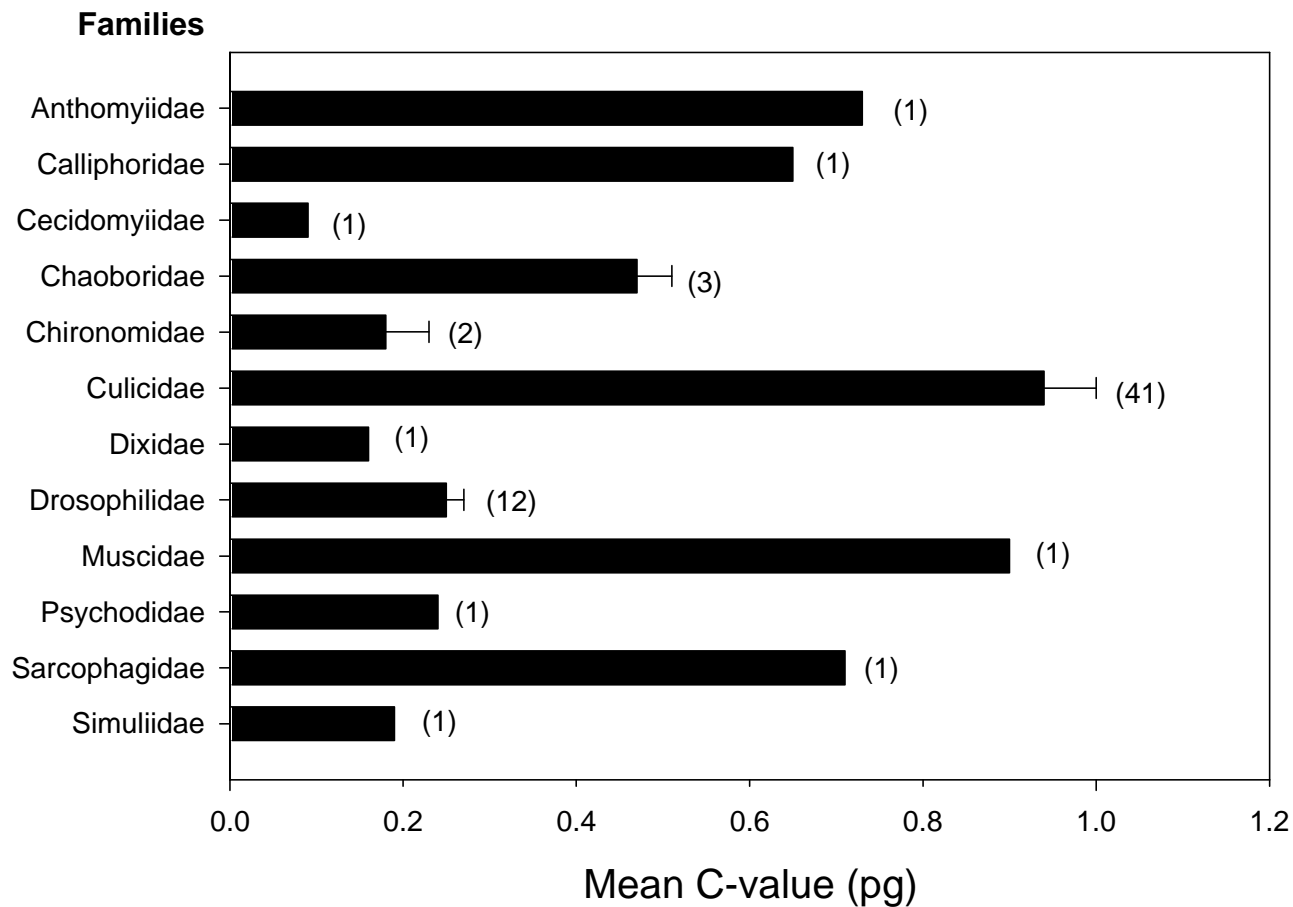


Figure 6.2. Summary of haploid genome size (C-value) variation in the 12 families of true flies (Order Diptera) studied to date. Families are arranged alphabetically. Numbers in brackets indicate the number of species measured per family. Error bars represent standard error. The mean C-value for the Diptera is $0.71\text{pg} \pm 0.05$ (range: 0.09-1.9pg), based on roughly 65 species, mostly fruit flies and mosquitoes. Genome sizes as in Appendix 11.

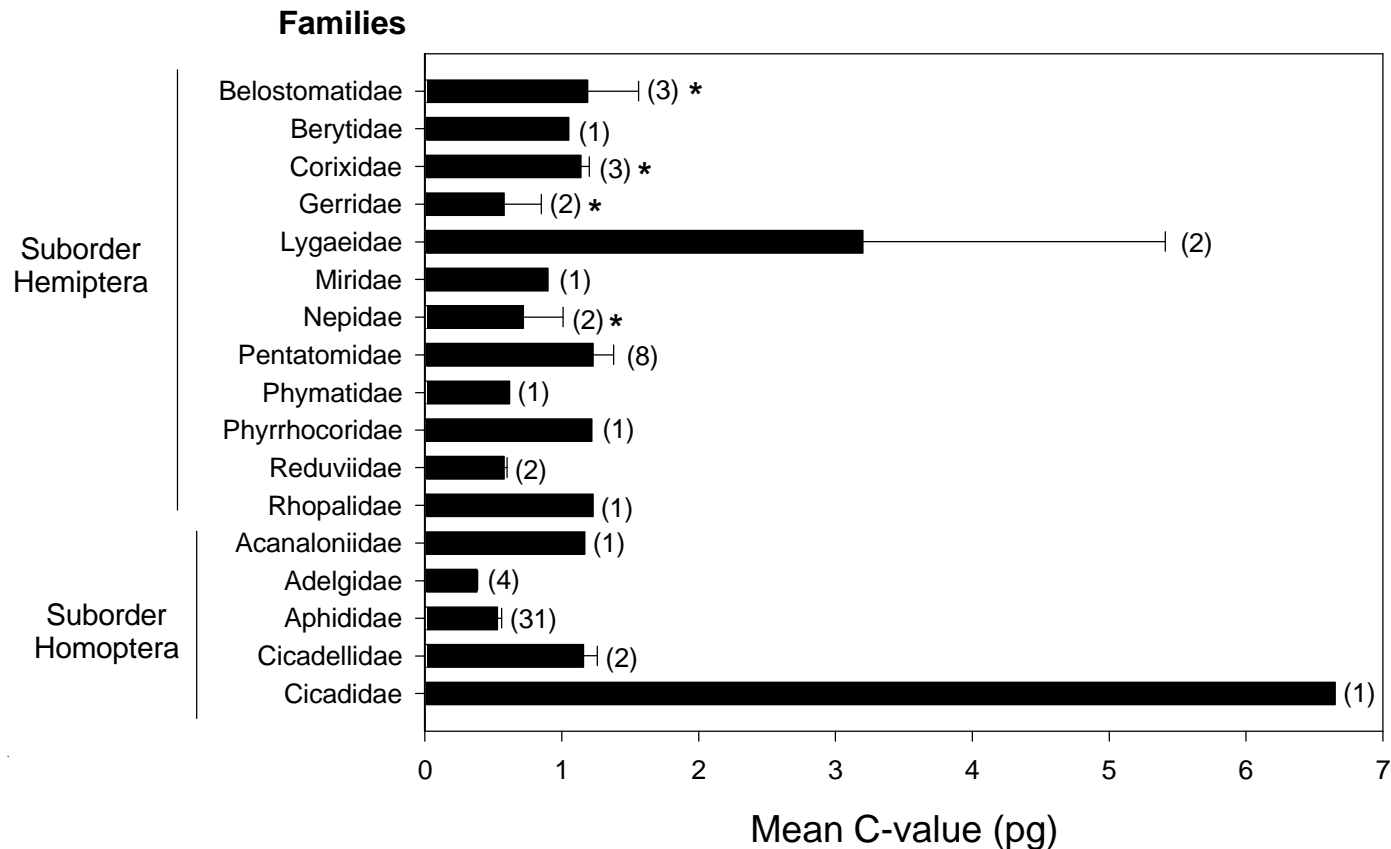


Figure 6.3. Summary of haploid genome size (C-value) variation in the 17 families of true bugs (Order Hemiptera) studied to date. Families are arranged alphabetically within suborders. Numbers in brackets indicate the number of species measured per family. Error bars represent standard error. The mean C-value for the Hemiptera as a whole is $0.91\text{pg} \pm 0.12$ (range: $0.18\text{-}6.7\text{pg}$), based on roughly 65 species. See the text for a discussion of variation within and among suborders. Aquatic families are marked with an asterisk. Genome sizes as in Appendices 11 and 12.

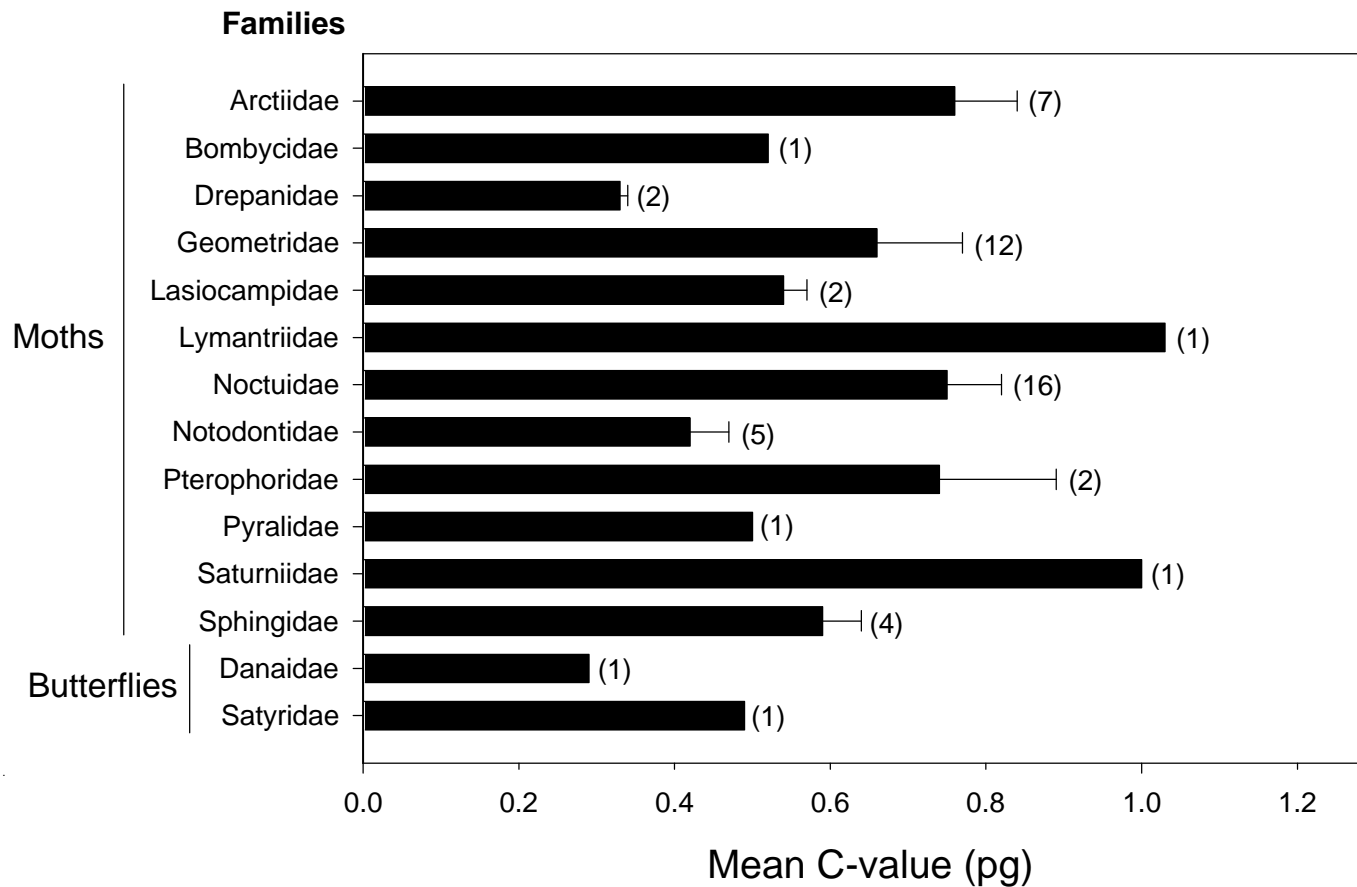


Figure 6.4. Summary of haploid genome size (C-value) variation in the 14 families of moths and butterflies (Order Lepidoptera) studied to date. Families are arranged alphabetically, although moths and butterflies are shown separately. Numbers in brackets indicate the number of species measured per family. Error bars represent standard error. The mean C-value for the Lepidoptera is $0.68\text{pg} \pm 0.04$ (range: 0.29-1.9pg), based on roughly 60 species, mostly macromoths. Genome sizes as in Appendices 11 and 12.

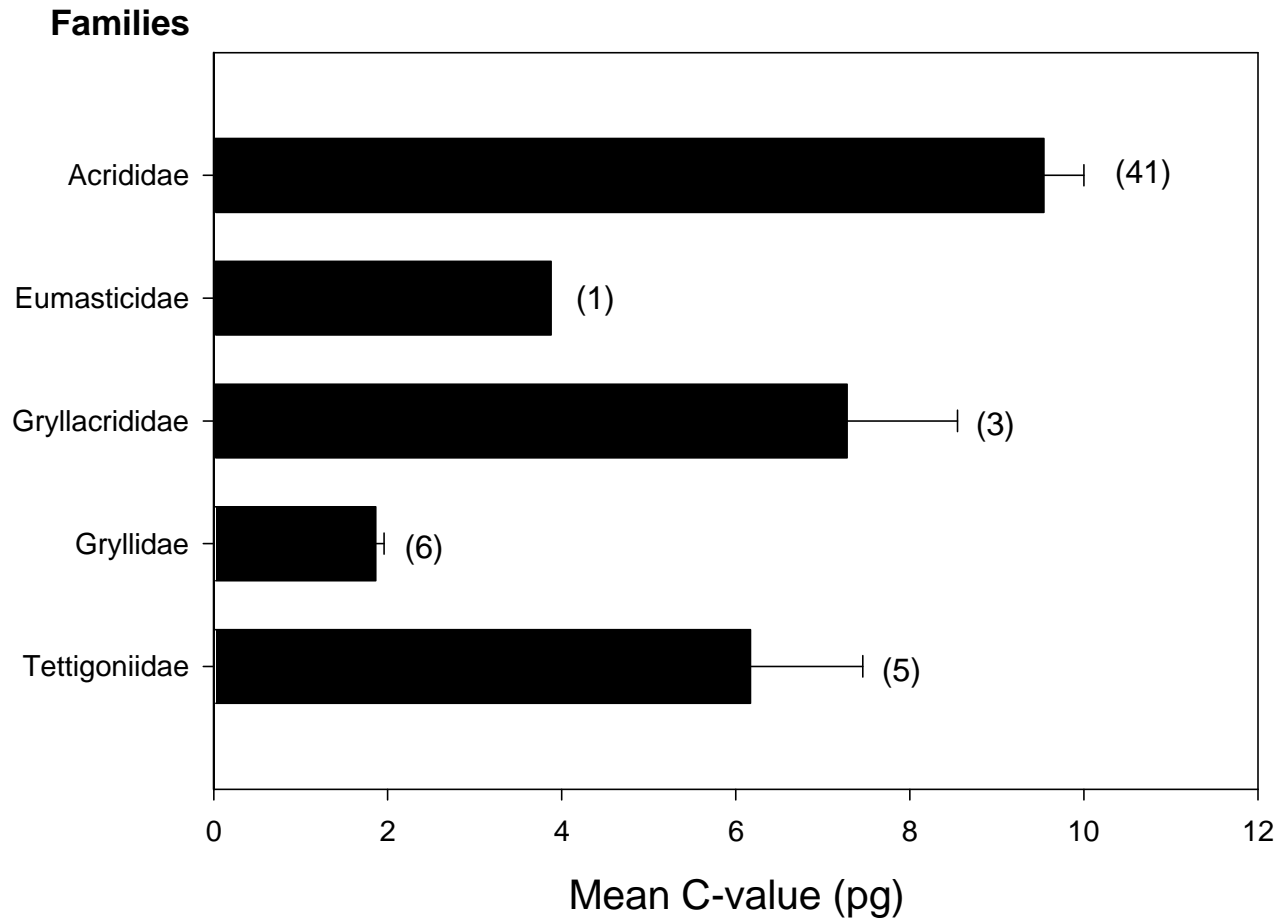


Figure 6.5. Summary of haploid genome size (C-value) variation in the 5 families of crickets, grasshoppers, and katydids (Order Orthoptera) studied to date. Families are arranged alphabetically. Numbers in brackets indicate the number of species measured per family. Error bars represent standard error. The mean C-value for the Orthoptera is $8.2\text{pg} \pm 0.5$ (range: 1.55-16.9pg), based on 56 species, mostly grasshoppers. Genome sizes as in Appendices 11 and 12.

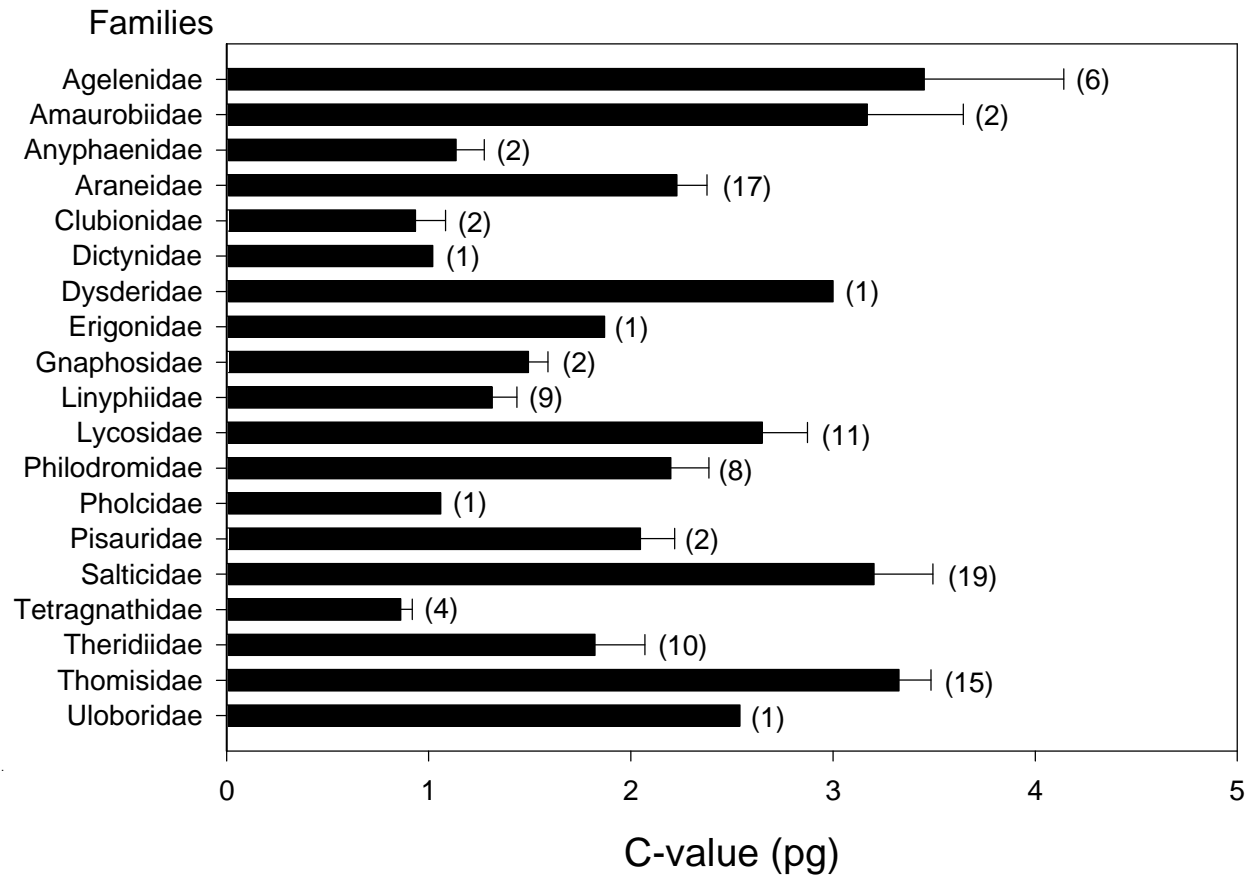


Figure 6.6. Summary of haploid genome size (C-value) variation in the 19 families of spiders examined in the present study. Numbers in brackets indicate the number of species measured per family. Error bars represent standard error. The mean C-value for these spiders is $2.4\text{pg} \pm 0.1$ (range: $0.74\text{-}5.7\text{pg}$), based on 114 species, all from the Infraorder Araneomorphae. Genome sizes as in Appendix 13.

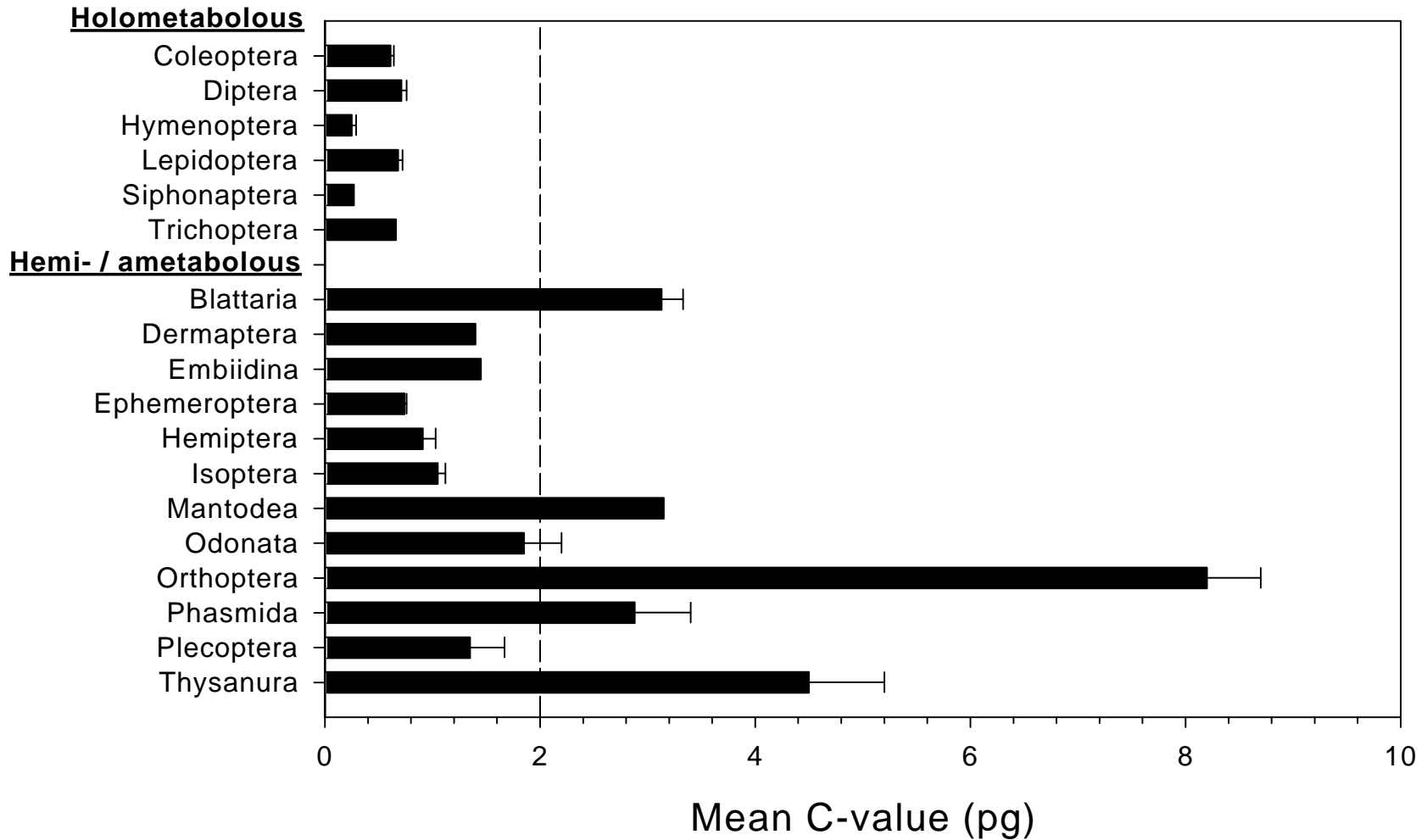


Figure 6.7. Mean genome sizes for the various orders of insects studied to date. The orders are grouped according to whether or not they undergo complete metamorphosis. The hypothetical 2pg threshold for complete metamorphosis is shown. Genome sizes as in Appendices 11 and 12.

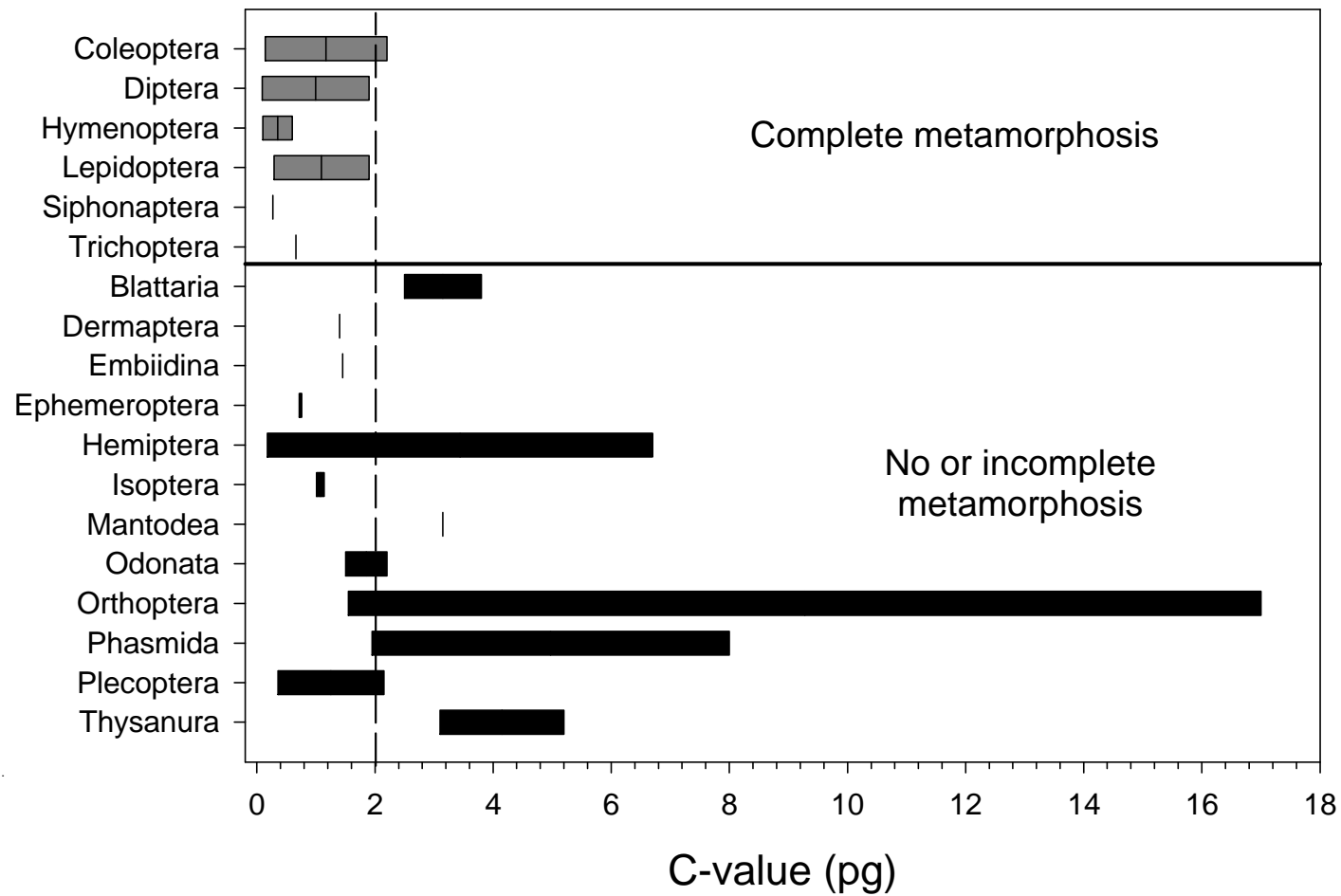


Figure 6.8. Ranges in genome size among the various orders of insects studied to date. The orders are grouped according to whether or not they undergo complete metamorphosis. The hypothetical 2pg threshold for complete metamorphosis is shown. Genome sizes as in Appendices 11 and 12.