TITLE: The evolution of DNA methylation and its relationship to sociality in insects

RUNNING TITLE: Social evolution and DNA methylation

Adam J. Bewick[1*], Kevin J. Vogel[2], Allen J. Moore[1], Robert J. Schmitz[1]

[1]Department of Genetics, University of Georgia, Athens, GA 30602, USA

[2]Department of Entomology, University of Georgia, Athens, GA 30602, USA

CO-CORRESPONDENCE: Adam J. Bewick (bewickaj@uga.edu) and Robert J. Schmitz (schmitz@uga.edu)

KEY WORDS: DNA methylation, whole genome bisulfite sequencing, social behavior, phylogenetic comparisons, molecular evolution

ABSTRACT

DNA methylation contributes to gene and transcriptional regulation in eukaryotes, and therefore has been hypothesized to facilitate the evolution of flexible traits such as sociality in insects. However, DNA methylation is sparsely studied in insects. Therefore, we documented patterns of DNA methylation across a wide diversity of insects. Furthermore, we tested the hypothesis that the DNA methylation system will be associated with presence/absence of sociality among insects. We also predicted that underlying enzymatic machinery is concordant with patterns of DNA methylation. We found DNA methylation to be widespread, detected in all orders examined except Diptera (flies). Whole genome bisulfite sequencing showed that orders differed in levels of DNA methylation. Hymenopteran (ants, bees, wasps and sawflies) had some of the lowest levels, including several potential losses. Blattodea (cockroaches) show all possible patterns, including a potential loss of DNA methylation in a eusocial species whereas solitary species had the highest levels. Phylogenetically corrected comparisons revealed no evidence that supports evolutionary dependency between sociality and DNA methylation. Species with DNA methylation do not always possess the typical enzymatic machinery. We identified a gene duplication event in the maintenance DNA methyltransferase 1 (DNMT1) that is shared by some hymenopteran, and paralogs have experienced divergent, non-neutral evolution. This diversity and non-neutral evolution of underlying machinery suggests alternative DNA methylation pathways may exist. Altogether, DNA methylation is highly variable in insects and is not a universal driver of social behavior. Future, functional studies are required to advance our understanding of DNA methylation in insects.

INTRODUCTION

DNA methylation is recognized as an important chromatin modification providing structural integrity and proper regulation of the genome for many species. In animals DNA methylation typically occurs at CG sites, established *de novo* by DNA methyltransferase 3 (DNMT3), and maintained by the maintenance methyltransferase DNMT1 (Kim et al. 2008; Goll and Bestor 2005; Cheng and Blumenthal 2008). Homologous to DNMT1 and 3 is DNMT2: a tRNA^{Asp} DNA

methyltransferase, which does not contribute to the DNA methylome (Goll et al. 2006). Despite CG DNA methylation conservation across the tree of life (Feng et al. 2010; Zemach et al. 2010; Huff and Zilberman 2014), the understanding of its contribution to genome function and complex traits is limited. Although it has been proposed that heritable epigenetic variation is an important contributor to evolutionary change, direct evidence for this hypothesis is sparse. In insects, DNA methylation is implicated in behavioral plasticity and social behavior, especially eusociality in Hymenoptera (Yan et al. 2014; Yan et al. 2015). At a mechanistic level, it is difficult to test the role of DNA methylation in social behavior because the generation of mutants depleted of DNA methylation are not available. As an alternative, comparative methods have been used to infer conserved function of DNA methylation (Elango et al. 2009; Bonasio et al. 2012). Comparisons between pairs of species within the order Hymenoptera (ants, bees, wasps and sawflies) support the association between DNA methylation and presence of non-reproductive castes (Elango et al. 2009; Bonasio et al. 2012; Patalano et al. 2013). Although others have observed little association between social behavior and DNA methylation (Libbrecht et al. 2016; Standage et al. 2016). However, comparative epigenomics in insects has been taxonomically piecemeal, and examining potential influences on insect sociality beyond eusocial traits have yet to be performed.

Here we test the hypothesis that DNA methylation is associated with the evolution of sociality, more broadly defined (Wilson 1971; Costa 2006) across insects and independent of phylogenetic history. We define sociality at as is commonly done (Wilson 1971; Costa 2006) – solitary (none), communal (group living without cooperation or care), subsocial (overlapping generations and parental care), and eusocial (overlapping generations, parental care, and reproductive division of labor). Because studies of DNA methylation in insects are currently taxonomically sparse, we first define the extent that DNA methylation occurs in insects. We tested for the presence of DNA methylation using whole genome bisulfite sequencing (WGBS) and utilized previously published data to generate a dataset comprising of 41 species of insects from six orders, including three orders where eusociality has evolved. To expand our taxonomic sampling and cover insect orders and species where sequencing has not yet been done, we documented the presence or absence of DNA methylation using $CpG_{O/E}$ in 123 species of insects from 11 orders. The

measure $CpG_{O/E}$ relies on natural, spontaneous demanination of methylated cytosines to thymines, and robustly recovers presence/absence of DNA methylation (see Materials and Methods). We then used both continuous (WGBS) and discrete ($CpG_{O/E}$) data in explicit phylogenetic comparative tests that take into account evolutionary history and non-independence of species. We tested our hypothesis at two broad levels of sociality; the evolution of any level of sociality (from solitary to eusocial) and the evolution of division of labor associated with eusociality. Finally, we examined the presence, distribution and evolution of DNA methyltransferases across the diversity of insects. We found widespread presence but diverse patterns of DNA methylation across the insect tree of life. Contrary to a prediction of a role for DNA methylation facilitating social evolution, we found no association between sociality and DNA methylation in insects. Instead we suggest that the presence of DNA methylation in a variety of insects suggests a much broader function of this chromatin modification. This is supported by the diverse patterns of presence/ absence of DNA methyltransferases, including duplication of DNMT1 in some groups and lack of DNMT3 in others. Functional studies are now needed to elucidate the role of DNA methylation in insects, and our work will facilitate identifying species suitable for such studies.

RESULTS AND DISCUSSION

DNA methylation levels are highly variable across insects. Estimates of DNA methylation from WGBS from 41 species from six orders (Blattodea [n=9], Hemiptera [n=5], Hymenoptera [n=12], Coleoptera [n=3], Lepidoptera [n=4], and Diptera [n=8]) revealed diverse levels of DNA methylation. DNA methylation was found in all insect orders except Diptera and in social and solitary species (Fig. 1; Table S1). DNA methylation levels within coding regions are higher than genome-wide levels for all species, which suggests a conserved pattern of DNA methylation across insects (Fig. 1; Table S1). Higher levels of DNA methylation within coding regions compared to genome-wide might partially reflect highly conserved protein coding genes in insects having the highest levels of DNA methylation (Glastad et al. 2014). DNA methylation levels within commonly studied hymenopteran are some of the lowest compared to other species sampled, and three species (*Aphidius ervi*, *Microplitis*

demolitor and *Microplitis mediator*) are identified as having completely absent or extremely reduced levels of DNA methylation (Fig. 1; Table S1). This loss or extreme reduction in DNA methylation is similar to what is observed in the eusocial species *Polistes dominula* (Hymenoptera) (Standage et al. 2016). Levels of DNA methylation genome-wide and within coding regions were overall higher in the order Blattodea (Fig. 1). Genome-wide and coding levels of DNA methylation were highest in the solitary cockroaches in the genus *Blattella*, *B. asahinai* and *B. germanica*, respectively. Furthermore, drastic differences in levels of DNA methylation were observed between the highly eusocial species *Reticulitermes flavipes* and *R. virginicus* within Blattodea (Fig. 1). Conversely, similar levels of DNA methylation were observed between closely related *Pemphigus spp.* of aphids (Hemiptera) with differing social behaviors (Fig. 1). Additionally, the solitary species *Acyrthosiphon pisum* had the highest levels of DNA methylation within in the Hemiptera. Overall, DNA methylation levels were variable in insects, with some species' levels reaching those observed in plants (Feng et al. 2010; Zemach et al. 2010; Niederhuth and Bewick et al. 2016; Takuno et al. 2016), where its contribution to complex traits still remains largely unknown.

DNA methylation is present across the insect tree of life. DNA methylation, signified by a bimodal distribution of $CpG_{O/E}$ values, is a robust estimator for the presence/absence of DNA methylation in insect genomes. When compared to estimates of DNA methylation from WGBS, $CpG_{O/E}$ accurately estimates presence/absence of DNA methylation in 18/18 (100%) species (see Materials and Methods, and Table S1). Therefore, to taxonomically expand on insect orders and species, the distribution of $CpG_{O/E}$ of 123 insect species from 11 orders were investigated for signatures of DNA methylation. Similarly to our WGBS data, DNA methylation was identified in insect species from all orders except Diptera (Fig. 2A and B). Species belonging to the order Diptera comprised ~46.3% (57/123) of the total number of species investigated, which was the largest number of species sampled from one order. The absence of bimodal distribution from members of both the Nematocera and Brachycera sub-orders suggests that DNA methylation was lost early on in dipteran evolution. The evolution of DNA methylation is deeply rooted in insects, going as far back as the Ordovician and the divergence of Diplura (~489.84-

436.68 MYA (Misof et al. 2014)), but is most likely older.

DNA methylation was found in insects exhibiting the complete range of social behavior based on modality of $CpG_{O/E}$ (Table S1). This includes insects classified as solitary, communal (group living), subsocial (parental care), and eusocial (cooperative brood care, overlapping generations and reproductive castes – references (Wilson 1971; Costa 2006)). For example, the solitary thrip *Frankliniella occidentalis* had an extremely pronounced bimodal $CpG_{O/E}$ distribution compared to the eusocial honey bee *Apis mellifera* (Fig. S1). The order Hymenoptera contains species that are solitary or eusocial; however, evidence for DNA methylation based on $CpG_{O/E}$ did not always co-occur with eusociality, and *vice versa* (Table S1). Moreover, all species examined within the order Lepidoptera are solitary, yet 8/10 showed evidence for DNA methylation based on the distribution of $CpG_{O/E}$, respectively (Table S1). Similarly, 3/4 solitary coleopteran species were expected to have DNA methylation based on bimodality of $CpG_{O/E}$. Interestingly, the six Diptera in the genus *Glossina* have sub-social behavior in the form of parental care (e.g., Table S1, the tsetse fly, *Glossina morsitans* morsitans (International Glossina Genome Initiative et al. 2014)), yet do not have DNA methylation based on the distribution of $CpG_{O/E}$. This is contrary to the burying beetle *Nicrophorus vespilloides*, which has DNA methylation based on WGBS and $CpG_{O/E}$ (Cunningham et al. 2015), and also provides parental care through provisioning developing offspring (Walling et al. 2008) (Table S1). Overall, based on modality of $CpG_{O/E}$ DNA methylation is found in all insect orders examined except Diptera.

Sociality and DNA methylation are not evolutionary dependent. DNA methylation has been proposed to control many aspects of sociality including behavior expressed in social interactions, caste determination, and learning and memory (reviewed in (Li-Byarlay and Hongmei 2016)). Other investigators have suggested that empirical evidence for caste-specific DNA methylation in social insects is weak (Libbrecht et al. 2016). However, relatedness of species has not been considered in previous studies, and is necessary to control for non-independence of species relationships (Felsenstein 1985). If sociality is dependent on DNA methylation we would expect a gain in DNA methylation to result in either a gain of social behavior or a predisposition to evolve social behavior. Thus, if sociality

is dependent on DNA methylation we would expect a phylogenetically-corrected correlation between these two traits. We do not find this. Sociality – including division of labor – is not evolutionary dependent on DNA methylation (Table 1). This lack of dependence holds true when DNA methylation is categorized as a discrete or continuous trait, and sociality is categorized as presence/absence or into multiple discrete classes (Fig. 1; Fig. 2A; Table 1). Although there are differences in transition rates between social behavior and DNA methylation states, they are not dependent on one another (Table 1; Fig. S2). For example, transitions from social to solitary occur more frequently than solitary to social when DNA methylation is fixed (Fig. S2). Also, transitions from solitary to social, *in lieu* of DNA methylation, occur more frequently than transitions from social to solitary (Fig. S2). Furthermore, based on Akaike information criterion (AIC), an independent model of evolution for each trait is preferred (Table 1). We further show that these traits are not dependent when 35 species within Hymenoptera are considered (Table 1). As dipterans composed a large fraction of species investigated in this study and do not have DNA methylation (Fig. 1 and 2), we were concerned that they may be biasing our analyses. However, when all species within Diptera are removed we similarly recover a lack of dependence between sociality and DNA methylation (Table 1). Together this evidence supports the hypothesis that sociality and DNA methylation are not evolutionary dependent.

Evolution of DNMT1, DNMT2 and DNMT3 and DNA methylation. DNMT1 is found in all orders of insects investigated in this study except Diptera (Fig. 3A; Table S2). Only DNMT2 is found in Diptera, which reflects the lack of DNA methylation in the genomes of all species belonging to this order (Fig. 2B; Table S1). The loss of DNA methylation most parsimoniously occurred once in all Diptera ~206.28-107.26 MYA (Misof et al. 2014), which likely coincided with the loss of DNMT1 and/or DNMT3.

Only DNMT1 and DNMT2 are found in lepidopteran species (Fig. 3A; Table S2). This observation supports the hypothesis that DNMT3 was lost from species belonging to this order between ~177.99-116.45 MYA (Misof et al. 2014). Also, the sister order to Lepidoptera, Trichoptera, also only possess DNMT1 and DNMT2. Therefore, the timing of this loss may be older, however this order is represented by only one species in our data. Despite missing DNMT3 from assembled transcriptomes or genomes – as in the lepidopteran *Bombyx mori* – evidence for

DNA methylation is still observed from WGBS (Table S1) data and from distributions of $CpG_{O/E}$ values (Fig. S1) and WGBS (Table S1) (Xiang et al. 2010). Similarly, 2/7 Coleoptera and 9/10 Hemiptera only possess DNMT1 and DNMT2, and DNA methylation is expected to be present given bimodality of $CpG_{O/E}$ (Table S1).

Several duplications were observed in the DNMT1 clade, but none coincide with presence or absence of DNA methylation or sociality. A duplication event shared by many hymenopteran gave rise to what is referred to as DNMT1a and DNMT1b (Fig. 4). Differences in codon and amino acid alignment tree topology suggest all hymenopterans or bees and ants shared the duplication, respectively (Fig. 4A; Fig S4). The amino acid alignment tree topology has higher node support than the codon alignment, which provides stronger support for the duplication event being shared by the superfamilies Apoidea (bees) and family Vespoidea (ants and some wasps) and places the timing for this duplication ~123-225 MYA (Ronquist et al. 2012) (Fig. 4A; Fig. S4). Species relationships suggest reciprocal losses among families within the superfamily Vespoidea: a loss of DNMT1a in Vespidae (some wasps) and a loss of DNMT1b in Formicidae (ants). Whereas DNMT1a and DNMT1b can both be found in the superfamily Apoidea (families Apidae, Halictidae, and Megachilidae). Chalcid wasps (family Chalcidoidea) experienced at least two superfamily-specific duplication events, as suggested by the monophyletic groups containing *Nasonia vitripennis* (Fig. 4A). Additional species-specific duplications – in-paralogs – were observed in orders Hemiptera and Lepidoptera (Fig. 3), but these could represent allelic variation of DNMT1.

Gene duplication is a source of genetic novelty and functional divergence, which can be characterized by non-neutral evolution (Lynch and Conery 2000). Divergent selection is observed between DNMT1a and DNMT1b in Apoidea (bees), with DNMT1a having experienced relaxed purifying selection and DNMT1b purifying selection (Fig. 4). Furthermore, positively selected sites were identified in DNMT1b, and these include one site (T432) in the CXXC zinc finger domain, and three (V575, A581, S919) in the BAH domain (Fig. 4). DNMT1 paralogs might maintain DNA methylation differently, including efficiency or rate, and spatially and/or temporally. The contribution of each paralog to the maintenance of the DNA methylome is unknown, and functional studies are required to address the fates of these duplicated genes.

DNMT3 is the most order-poor of the DNA methyltransferases, and was only identified in species belonging to Blattodea, Coleoptera, Hemiptera, and Hymenoptera (Fig. 3A; Table S2). For species with annotated genomes, DNMT3 is always accompanied by the presence of DNMT1 (Table S2). Given the scarcity of DNMT3 in insects and positive association between the presence and levels of DNA methylation and DNMT1 only, suggests DNMT3 may be dispensable for DNA methylation or DNMT1 compensates for DNMT3. However, DNMT1 is missing protein domains typically associated with metazoan *de novo* DNA methyltransferases. For example, DNMT1 does not contain a PWWP domain (PF00855), which interacts with DNA and histone lysine modified nucleosomes (Qiu et al. 2010; Qin et al. 2014). PWWP domain containing proteins exist (e.g., *Bombyx mori*), however none of these proteins contain a C-5 cytosine-specific DNA methylase domain (PF00145), which is required for DNA methylation (Bestor and Verdine 1994; Goll et al. 2005; Cheng et al. 2008). Based on these observations three possible alternative mechanisms explain the presence of DNA methylation *in lieu* of typical enzymes: (i) *de novo* DNA methylation does not occur via a DNMT-like protein; (ii) DNA methylation is not reprogrammed during embryogenesis and is robustly maintained by DNMT1 during each cell replication; or (iii) as has been shown in a plant species (Bewick, Li, Niederhuth, Willing et al. 2016; Bewick et al. 2016), non-neutral evolution could have occurred at DNA methyltransferases, which affects their ability to methylate or maintain methylation. However, alternative mechanisms might exist, and functional tests of DNA methylation mechanisms in insects warrant further investigation.

CONCLUSION

We have conducted the largest phylogenetic investigation of DNA methylation in insects. Our investigation suggests that DNA methylation is too variable in insects to support a single, common role in regulating traits expressed in social interactions. DNA methylation is common in insects, but the level of DNA methylation is fluid and is not more likely to be associated with the presence of sociality. Although this does not preclude a role for DNA methylation in regulating caste in eusocial insects, one possibility is that DNA methylation in most insects that display sociality is a downstream consequence of upstream environment, genetic or hormonal factors

(Kapheim et al. 2015; Wallberg et al. 2016). Furthermore, the gain, loss, and duplication of DNMT1 suggests that there may be novel functions and justifies further studies using RNAi or CRISPR/Cas9 to mutate DNA methyltransferases (Gilles et al. 2015). For now, the universal role of DNA methylation in insects remains elusive, and is likely to remain so until functional tests are made to advance our understanding of this evolutionary conserved and important chromatin modification.

MATERIALS AND METHODS

Tissue collection and DNA extraction. All samples were collected from established laboratory colonies with the following exceptions. *Cryptocercus garciai* was collected from northeast Georgia by Brian Forschler, *Polistes carolina* were collected from the campus of the University of Georgia by K. J. Vogel.

DNA extractions were performed on tissues from either freshly sacrificed insects or ethanol preserved samples. For samples likely to contain significant gut microbial contamination (Blattodea), guts were dissected and discarded. Approximately 10 mg of material was frozen in liquid nitrogen, ground with a pestle. Samples were then extracted with the Qiagen DNEasy Mini kit following the manufacturer's instructions for animal tissues.

Behavior classification. Insect species were classified into one of four social behavior categories based on the presence or absence of well-defined traits (Wilson 1971; Costa 2006). *Eusocial* insects must share a common nest site, exhibit cooperative rearing of young, reproductive division of labor with sterile or less fecund workers, and have overlap of generations. Species that exhibited only parental care or cooperative brood rearing were classified as *subsocial*. Species with individuals of the same generation that use the same composite nest site and do not cooperate in brood care were classified as *communal*.

Whole-genome bisulfite sequencing and levels of DNA methylation. MethylC-seq libraries were prepared according to the following protocol (Urich et al. 2015). Sequencing data for *Acyrthosiphon pisum*, *Aedes aegypti*, *Aedes albopictus*, *Anopheles gambiae*, *Apis mellifera*, *Blattella germanica*, *Camponotus floridanus*,

Copidosoma floridanum, *Culex quinquefasciatus*, *Dinoponera quadriceps*, *Drosophila melanogaster*, *Harpegnathos saltator*, *Microplitis demolitor*, *Nasonia vitripennis*, *Nicrophorus vespilloides*, *Polistes canadensis*, *Solenopsis invicta*, *Tribolium castaneum* was aligned to their respective genome assembly using the methylpy pipeline (Schultz et al. 2015) (Table S1). *Blattella asahinai* was aligned to its sister species *Blattella germanica*. Similarly, *Microplitis mediator* was aligned to its sister species *Microplitis demolitor*. *Dinoponera quadriceps* was removed from Fig. 1 and associated analyses because of possible genomic sequence contamination with other insect species. Species for MethylC-seq were chosen to maximize taxonomic breadth based on previously published insect DNA methylomes. Thus, species within the same genus with published methylomes were often limited to a single representative. However, *a priori* knowledge of closely related species with differences in social behavior was taken into consideration. Overall, the species chosen for MethylC-seq represent phylogenetically deep and shallow switches for social behavior. Other species, including *Ceratina calcarata, Locusta migratoria* and *Schistocerca gregaria*, were excluded because no genome annotation exists on GenBank.

In brief, reads were trimmed of sequencing adapters using Cutadapt (Martin and Marcel 2011), and then mapped to both a converted forward strand (cytosines to thymines) and converted reverse strand (guanines to adenines) using bowtie (Langmead et al. 2009). Reads that mapped to multiple locations, and clonal reads were removed. Weighted DNA methylation was calculated for each sequence context (CG, CHG and CHH) by dividing the total number of aligned methylated reads by the total number of methylated plus un-methylated reads. Slight differences in levels of DNA methylation can be found between this and previously published studies (e.g., Bonasio et al. 2012). These differences are shared by all studies and are most likely due to technical and methodological differences. The majority of these differences are negated within this study due to the use of identical methods applied to all samples, including mapping of MethylC-Seq reads, categorization of coding sequences, and estimates of DNA methylation levels. Thus, levels of DNA methylation are comparable within this study. Furthermore, the discrete categorization of DNA methylation is not affected by slight differences between different social castes or individuals of a population, as genome-wide reductions or gains in DNA methylation are not observed (Elango et al. 2009). For the remaining

species, genome-wide levels of DNA methylation were estimated using $FAST^mC$ and the --animal model (Bewick et al. 2015).

CG DNA methylation in insects is enriched in coding sequences (see Lyko et al. 2010; Beeler et al. 2014) and a strong correlation between genome-wide and within coding sequence levels of CG DNA methylation is observed (Fig. S3). Therefore, using genome-wide levels of CG DNA methylation estimated from $FAST^mC$, we were able to extrapolate levels of CG DNA methylation within coding sequence for species without sequenced genomes (Table S1).

Measurement of $CpG_{O/E}$ and tests for bimodality. $CpG_{O/E}$ is a metric of CpG dinucleotides normalized by G and C nucleotide content (GC content) and length (bp) of a specific region of interest (e.g., a transcript or protein coding gene) (Patalano et al. 2015). Due to spontaneous deamination of methylated cytosines, genes that are hypermethylated are expected to have a lower $CpG_{O/E}$ value than hypomethylated genes. Thus, in a mixture of genes that are methylated and low to un-methylated, a bimodal distribution of $CpG_{O/E}$ values is expected. Conversely, a unimodal distribution is suggestive of a set of genes that are mostly low to un-methylated. We therefore used this metric to expand our sampling of insect species for the presence of DNA methylation in 11 orders: Odonata (n=1), Ephemeroptera (n=1), Blattodea (n=2), Thysanoptera (n=1), Hemiptera (n=9), Phthiraptera (n=1), Hymenoptera (n=32), Coleoptera (n=7), Trichoptera (n=1), Lepidoptera (n=11), and Diptera (n=57). The $CpG_{O/E}$ value for each gene within 125 total transcriptomes or gene annotations was defined as:

$$CpG_{O/E} = \left(\frac{l^2}{l}\right) \times \left(\frac{P_{CpG}}{P_C \times P_G}\right) \tag{1}$$

where P_{CpG}, P_C, and P_G are the frequencies of CpG dinucleotides, C nucleotides, and G nucleotides, respectively, estimated from each gene of length (l) in bp. Only exonic sequences of a gene were considered when estimating $CpG_{O/E}$.

The modality of $CpG_{O/E}$ distributions was tested using Gaussian mixture modeling (mclust v5.2). Two modes were modeled for each $CpG_{O/E}$ distribution, and the subsequent means and 95% confidence interval (CI) of the means were compared with overlapping or non-overlapping CI's signifying unimodality or

bimodality, respectively. Based on the largest set of genomes and WGBS data to date, Gaussian mixture modeling using mclust identified 18/18 (100%) species correctly for presence/absence of DNA methylation within coding regions. Thus, $CpG_{O/E}$ modality tested with Gaussian mixture modeling is a robust and accurate predictor of DNA methylation.

Phylogenetic comparative methods. The R package phytools (Revell 2011) was used for all phylogenetic comparative methods. Two tests of correlated evolution were conducted. (i) Pagel's (Pagel 1994) method for detecting correlated evolution of two binary traits. Briefly, Pagel's method uses a continuous-time Markov model to simultaneously estimate transition rates in pairs of binary characters on a phylogeny. These rates are then used to test whether an independent or dependent model of evolution is preferred using the likelihood ratio test (LRT). DNA methylation was categorized into un-methylated or methylated based on the distribution of $CpG_{O/E}$, unimodal or bimodal, respectively. Sociality was categorized into social (eusocial, subsocial, and communal) or solitary binary traits. Similarly, the *division of labor* trait was categorized into species with (caste$^+$) or without (caste$^-$) castes. (ii) Phylogenetic generalized least squares (PGLS) was used to correlate continuously categorized estimates of DNA methylation generated through WGBS and alignment to reference assemblies (methylpy), or non-referenced based methods (*FAST^{m}C*) to discretely coded social traits (eusocial, subsocial, communal, and solitary). BEAST 2 (Bouckaert et al. 2014) was used to estimate a multilocus coalescent tree, which was used to control for relatedness of species (non-independence) for both comparative tests. For (i) the phylogenetic tree was estimated from a subset of previously identified orthologous protein coding loci (58 of the >1400 described in Misof et al. 2014). These loci were chosen because each contained no more than 5% missing species for 123 insects, and outgroups *Catajapyx aquilonaris* (Diplura) and *Daphnia pulex* (Crustacea) based on best BLASTp hit to a set of core species (*Acyrthosiphon pisum*, *Anopheles gambiae*, *Acromyrmex echinatior*, *Bombyx mori*, *Drosophila melanogaster*, *Apis mellifera*, *Ixodes scapularis*, *Nasonia vitripennis*, *Pediculus humanus*, *Tribolium castaneum*, *Zootermopsis nevadensis*, *Daphnia pulex*). Loci were aligned using PASTA (Mirarab et al. 2015), and Gblocks was used to identify sections of conserved protein coding sequence (Castresana et al. 2000). For (ii) a

smaller number of loci were used due to the obscurity of species and available data on Genbank. Loci used for both trees can be found in Table S3. For both trees, the program Tracer (http://tree.bio.ed.ac.uk/software/tracer/) was used to assess stationarity and effective sample size (ESS; ≥100) of the Markov Chain Monte Carlo (MCMC) chains.

DNA methyltransferase (DNMT) phylogeny and evolution. DNA methyltransferases (DNMT1, 2, and 3) were curated from 125 insect species through homology searches using BLASTp and previously identified DNMT1, 2, and 3 proteins in *Apis mellifera*, *Bombyx mori* (DNMT1 only), *Drosophila melanogaster* (DNMT2 only), and *Nicrophorus vespilloides*. A series of alignment and phylogenetic estimation steps were conducted eliminate partial and poor sequences, which can affect alignment and subsequently the topology of phylogeny. Protein sequences were aligned using the program PASTA, and back-translated using the CDS sequence to generate an in-frame codon alignment. RAxML with 1000 rapid bootstrap replicates and the GTR+G model of nucleotide substitution was used to generate the phylogeny (Stamatakis 2014). The phylogeny was rooted to the DNMT2 clade in the program FigTree (http://tree.bio.ed.ac.uk/software/figtree/) and exported for stylization. An identical method was used to generate the Hymenoptera-specific DNMT1 gene tree. Nine iterations of alignment and phylogeny construction were performed removing sequences on long branches, which is indicative of low sequence homology. For species with assembled transcriptomes or gene annotations, additional DNA methyltransferases were identified by Interproscan (Jones et al. 2014), and filtering those sequences with a DNA methylase domain (PF00145). These DNA methylase domain-containing sequences were then subjected to the previously mentioned homology searches using BLASTp.

Codon analysis. Similar methodology as described above was used to construct phylogenetic trees for testing hypotheses on the rates of evolution in a phylogenetic context. However, the program Gblocks was used to identify conserved amino acids in codon alignments. The parameters for Gblocks were kept at the default settings, except we allowed for %50 gapped positions. The program Phylogenetic Analysis by Maximum Likelihood (PAML) was used to test branches (branch test) and sites along branches (branch-site test) for deviations from the background rate of molecular

evolution (*dN/dS*; ω) and for deviations from the neutral expectation, respectively (Yang 2007). Positively selected sites were determined by Bayes Empirical Bayes (BEB) score of ≥0.95. Branches tested and a summary of each test can be found in Table S4.

ACKNOWLEDGEMENTS

AUTHOR CONTRIBUTIONS

The study was conceived by AJB and RJS. All authors were involved in the experimental design. KJV and AJM curated samples. KJV and AJM classified social behavior for insects used in this study. AJB performed all analyses, with contribution from KJV to the phylogenetic analyses. AJB wrote the manuscript with contributions made by all authors.

FIGURES AND TABLES

Figure 1. WGBS reveals extensive variation of DNA methylation in insects. (A) Genomic levels of DNA methylation in insects ranges from zero (all Diptera examined) to ~14% (*Blattella asahinai*). Higher ranges are observed for coding regions; zero to ~42% (*Blattella germanica*). Overall, levels are highest in Blattodea, and do not always associate with social species. (B) A species tree constructed from

nuclear and mitochondrial loci, which was used in Phylogenetic Generalized Least Squares (PGLS) analysis. Results from this analysis revealed that there is no correlation between social behavior and DNA methylation (Table 1). Values at nodes are posterior probabilities <0.95; all blank nodes have ≥0.95 posterior probability.

Figure 2. DNA methylation in insects is not always associated with social behavior. (A) Relationships of 123 insect species, and the outgroups *Catajapyx aquilonaris* (Dipluran) and *Daphnia pulex* (Crustacea) investigated with sociality, division of labor and DNA methylation scored as a binary (presence/absence) trait. The tree was constructed from 58 nuclear protein coding loci, and was used in Pagel's test for evolutionary dependence. All nodes except one had posterior probability of 1.00, which is indicated on the phylogeny. (B) A chronogram of insect order relationships with sociality, division of labor and DNA methylation scored as binary (presence/absence) trait. The chronogram was modified from Misof et al. 2014. For (A) and (B) traits are represented as shaded boxes above each species or order. Half-filled boxes indicate the trait is variable within the corresponding order.

Figure 3. Evolution of DNMT1, 2 and 3 across insects. (A) Relationships of DNMT1, 2, and 3 in insects, Diplura and Ixodida (Arachnida). DNMT2 can be found in all insect orders investigated, while DNMT1 and 3 are more order-poor. The insect order for each sequence is provided in square brackets following the GenBank or genome annotation accession number. The tree was rooted to DNMT2.

Figure 4. Divergent non-neutral evolution of DNMT1a and b in Apoidea. (A) Hypothesized relationships among DNMT1 in Hymenoptera suggests a duplication event shared by the superfamilies Apoidea (bees) and Vespoidea (ants and some wasps) gave rise to what is referred to as DNMT1a and b (Fig. S4). DNMT1a appears to have been lost from Vespidae (some wasps) and DNMT1b from Fromicidae (ants), whereas both DNMT1a and DNMT1b were retained in Apoidea (families Apidae, Halictidae, and Megachilidae). Divergent selection between DNMT1a and DNMT1b in Apoidea suggests the former is under relaxed purifying selection and the latter is under purifying selection. *Bombyx mori* (Lepidoptera) was used to root the tree, and was excluded from PAML analyses. (B) Several sites in DNMT1b were identified as under positive selection (yellow circles), with one site

(T432) in the CXXC zinc finger domain, and three (V575, A581, S919) in the BAH domain. The crystal structure of *Apis mellifera* DNMT1b was predicted from *Mus musculus* DNMT1. Details including red spheres, dark and light blue colouring specify Zn^{2+} ions, CXXC domain and s-adenosyl-l-homocysteine, the secondary product from the DNA methylation reaction it performs, according to *Mus musculus* DNMT1, respectively. Bootstrap support values are characterized as shaded circles. *dN/dS* (ω) values are for the most preferred branch model.

Table 1. Sociality and DNA methylation are not evolutionary dependent. Upper half: Output from Pagel's method for evolutionary dependence (implemented in phytools[47]) using the phylogeny and traits from Fig. 2A. The dependent model of trait evolution is not preferred based on the likelihood ratio test (LRT), and the independent model is preferred based on Akaike information criterion (AIC). Lower half: Output from Phylogenetic Generalized Least Squares (PGLS) implemented in phytools) using the phylogeny and traits from Fig. 1. A model of zero (0) phylogenetic signal (λ) and evolution under Brownian motion was preferred over a model where λ was estimated by maximum likelihood. For both tests, P-values represent the significance of correlations between trait *x* and *y*. P-values for PGLS are given for the preferred model.

REFERENCES

Beeler SM, Wong GT, Zheng JM, Bush EC, Remnant EJ, Oldroyd BP, Drewell RA. 2014. Whole-genome DNA methylation profile of the jewel wasp (*Nasonia vitripennis*). G3 4:383–388.

Bestor TH, Verdine GL. 1994. DNA methyltransferases. Curr Opin Cell Biol. 6:380–389.

Bewick AJ, Hofmeister BT, Lee K, Zhang X, Hall DW, Schmitz RJ. 2015. FASTmC: a suite of predictive models for non-reference-based estimations of DNA methylation. G3 6:447–452.

Bewick AJ, Ji L, Niederhuth CE, Willing EM, Hofmeister BT, Shi X, Wang L, Lu Z, Rohr NA, Hartwig B, et al. 2016. On the origin and evolutionary consequences of gene body DNA methylation. Proc Natl Acad Sci USA. 113:9111–9116.

Bewick AJ, Niederhuth CE, Rohr NA, Griffin PT, Leebens-Mack J, Schmitz RJ. 2016. The evolution of CHROMOMETHYLASES and gene body DNA methylation in plants. biorxiv doi:10.1101/054924.

Bonasio R, Li Q, Lian J, Mutti NS, Jin L, Zhao H, Zhang P, Wen P, Xiang H, Ding Y, et al. 2012. Genome-wide and caste-specific DNA methylomes of the ants Camponotus floridanus and Harpegnathos saltator. Curr Biol. 22:1755–1764.

Bouckaert R, Heled J, Kühnert D, Vaughan T, Wu CH, Xie D, Suchard MA, Rambaut A, Drummond AJ. 2014.BEAST 2: a software platform for Bayesian evolutionary analysis. PLoS Comput Biol. 10:e1003537.

Castresana J. 2000. Selection of conserved blocks from multiple alignments for their use in phylogenetic analysis. Mol Biol Evol. 17:540–552.

Cheng X, Blumenthal RM. 2008. Mammalian DNA methyltransferases: a structural perspective. Structure 16:341–350.

Costa JT. 2006. The Other Insect Societies. Harvard University Press.

Cunningham CB, Ji L, Wiberg RA, Shelton J, McKinney EC, Parker DJ, Meagher RB, Benowitz KM, Roy-Zokan EM, Ritchie MG, et al. 2015. The genome and methylome of a beetle with complex social behavior, *Nicrophorus vespilloides* (Coleoptera: Silphidae). Genome Biol Evol. 7:3383–3396.

Elango N, Hunt BG, Goodisman MAD, Yi SV. 2009. DNA methylation is widespread and associated with differential gene expression in castes of the honeybee, *Apis mellifera*. Proc Natl Acad Sci USA. 106:11206–11211.

Felsenstein J. 1985. Phylogenies and the comparative method. Am Nat. 125:1–15.

Feng S, Cokus SJ, Zhang X, Chen PY, Bostick M, Goll MG, Hetzel J, Jain J, Strauss SH, Halpern ME, et al. 2010. Conservation and divergence of methylation patterning in plants and animals. Proc Natl Acad Sci USA. 107:8689–8694.

Gilles AF, Schinko JB, Averof M. 2015. Efficient CRISPR-mediated gene targeting and transgene replacement in the beetle *Tribolium castaneum*. Development 142:2832-2839.

Glastad KM, Hunt BG, Goodisman MAD. 2014. Evolutionary insights into DNA methylation in insects. Curr Opin Insect Sci. 1:25–30.

Goll MG, Bestor TH. 2005. Eukaryotic cytosine methyltransferases. Annu Rev Biochem. 74:481–514.

Goll MG, Kirpekar F, Maggert KA, Yoder JA, Hsieh CL, Zhang X, Golic KG, Jacobsen SE, Bestor TH. 2006. Methylation of tRNA^{Asp} by the DNA methyltransferase homolog Dnmt2. Science 311:395–398.

Huff JT, Zilberman D. 2014. Dnmt1-independent CG methylation contributes to nucleosome positioning in diverse eukaryotes. Cell 156:1286–1297.

International Glossina Genome Initiative. 2014. Genome sequence of the tsetse fly (*Glossina morsitans*): vector of African trypanosomiasis. Science 344:380–386.

Jones P, Binns D, Chang HY, Fraser M, Li W, McAnulla C, McWilliam H, Maslen J, Mitchell A, Nuka G, et al. 2014. InterProScan 5: genome-scale protein function classification. Bioinformatics 30:1236–1240.

Kapheim KM, Pan H, Li C, Salzberg SL, Puiu D, Magoc T, Robertson HM, Hudson ME, Venkat A, Fischman BJ, et al. 2015. Social evolution. Genomic signatures of evolutionary transitions from solitary to group living. Science 348:1139-1143.

Kim JK, Samaranayake M, Pradhan S. 2008. Epigenetic mechanisms in mammals. Cell Mol Life Sci. 66:596–612.

Kocher SD, Li C, Yang W, Tan H, Yi SV, Yang X, Hoekstra HE, Zhang G, Pierce NE, Yu DW. 2013. The draft genome of a socially polymorphic halictid bee, *Lasioglossum albipes*. Genome Biol. 14:R142.

Langmead B, Trapnell C, Salzberg SL. 2009. Ultrafast and memory-efficient alignment of short DNA sequences to the human genome. Genome Biol. 10:doi: 10.1186/gb-2009-10-3-r25.

Li-Byarlay H, Hongmei L-B. 2016. The function of DNA methylation marks in social insects. Front Ecol Evol. 4.

Libbrecht R, Oxley PO, Keller L, Kronauer DJC. 2016. Robust DNA methylation in the clonal raider ant brain. Curr Biol. 26:391–395.

Lyko F, Foret S, Kucharski R, Wolf S, Falckenhayn C, Maleszka R. 2010. The honey bee epigenomes: differential methylation of brain DNA in queens and workers. PLoS Biol. 8:e1000506.

Lynch M, Conery JS. 2000. The evolutionary fate and consequences of duplicate genes. Science. 90:1151–1155.

Martin M, Marcel M. 2011. Cutadapt removes adapter sequences from high-throughput sequencing reads. EMBnet.journal 17:10.

Mirarab S, Nguyen N, Guo S, Wang LS, Kim J, Warnow T. 2015. PASTA: ultra-large multiple sequence alignment for nucleotide and amino-acid sequences. J Comput Biol. 22:377–386.

Misof B, Liu S, Meusemann K, Peters RS, Donath A, Mayer C, Frandsen PB, Ware J, Flouri T, Beutel RG, et al. 2014. Phylogenomics resolves the timing and pattern of insect evolution. Science 346:763–767.

Niederhuth CE, Bewick AJ, Ji L, Alabady M, Kim KD, Page JT, Li Q, Rohr NA, Rambani A, Burke JM, et al. 2016. Widespread natural variation of DNA methylation within angiosperms. doi:10.1101/045880.

Pagel M. 1994. Detecting correlated evolution on phylogenies: a general method for the comparative analysis of discrete characters. Proc R Soc Lond [Biol] 255:37–45.

Patalano S, Vlasova A, Wyatt C, Ewels P, Camara F, Ferreira PG, Asher CL, Jurkowski TP, Segonds-Pichon A, Bachman M, et al. 2015. Molecular signatures of plastic phenotypes in two eusocial insect species with simple societies. Proc Natl Acad Sci USA. 112:13970–13975.

Qin S, Su Q, Jinrong M. 2014. Structure and function of the nucleosome-binding PWWP domain. Trends Biochem Sci. 39:536–547.

Qiu C, Sawada K, Zhang X, Cheng X. 2002. The PWWP domain of mammalian DNA methyltransferase Dnmt3b defines a new family of DNA-binding folds. Nat Struct Biol. 9:217–224.

Revell LJ. 2011. phytools: an R package for phylogenetic comparative biology (and other things). Methods Ecol Evol. 3:217–223.

Ronquist F, Klopfstein S, Vilhelmsen L, Schulmeister S, Murray DL, Rasnitsyn AP. 2012. A total-evidence approach to dating with fossils, applied to the early radiation of the hymenoptera. Syst Biol. 61:973–999.

Schultz MD, He Y, Whitaker JW, Hariharan M, Mukamel EA, Leung D, Rajagopal N, Nery JR, Urich MA, Chen H, et al. 2015. Human body epigenome maps reveal noncanonical DNA methylation variation. Nature 523:212–216.

Stamatakis A. 2014. RAxML version 8: a tool for phylogenetic analysis and post-analysis of large phylogenies. Bioinformatics 30:1312–1313.

Standage DS, Berens AJ, Glastad KM, Severin AJ, Brendel VP, Toth AL. 2016. Genome, transcriptome and methylome sequencing of a primitively eusocial wasp reveal a greatly reduced DNA methylation system in a social insect. Mol Ecol. 25:1769-1784.

Takuno S, Shohei T, Jin-Hua R, Gaut BS. 2016. Evolutionary patterns of genic DNA methylation vary across land plants. Nat Plants 2:15222.

Urich MA, Nery JR, Lister R, Schmitz RJ, Ecker JR. 2015. MethylC-seq library preparation for base-resolution whole-genome bisulfite sequencing. Nat Protoc. 10:475–483.

Wallberg A, Pirk CW, Allsopp MH, Webster MT. 2016. Identification of multiple loci associated with social parasitism in honeybees. PLoS Genet. 12:e1006097.

Walling CA, Stamper CE, Smiseth PT, Moore AJ. 2008. The quantitative genetics of sex differences in parenting. Proc Natl Acad Sc. USA. 105:18430–18435.

Wilson EO. 1971. The Insect Societies. Belknap Press.

Xiang H, Zhu J, Chen Q, Dai F, Li X, Li M, Zhang H, Zhang G, Li D, Dong Y, et al. 2010. Single base–resolution methylome of the silkworm reveals a sparse epigenomic map. Nat Biotechnol. 28:516–520.

Yan H, Bonasio R, Simola DF, Liebig J, Berger SL, Reinberg D. 2015. DNA methylation in social insects: how epigenetics can control behavior and longevity. Annu Rev Entomol. 60:435–452.

Yan H, Simola DF, Bonasio R, Liebig J, Berger SL, Reinberg D. 2014. Eusocial insects as emerging models for behavioural epigenetics. Nat Rev Genet. 15:677–688.

Yang Z. 2007. PAML 4: phylogenetic analysis by maximum likelihood. Mol Biol Evol. 24:1586-1591.

Zemach A, McDaniel IE, Silva P, Zilberman D. 2010. Genome-wide evolutionary analysis of eukaryotic DNA methylation. Science 328:916–919.
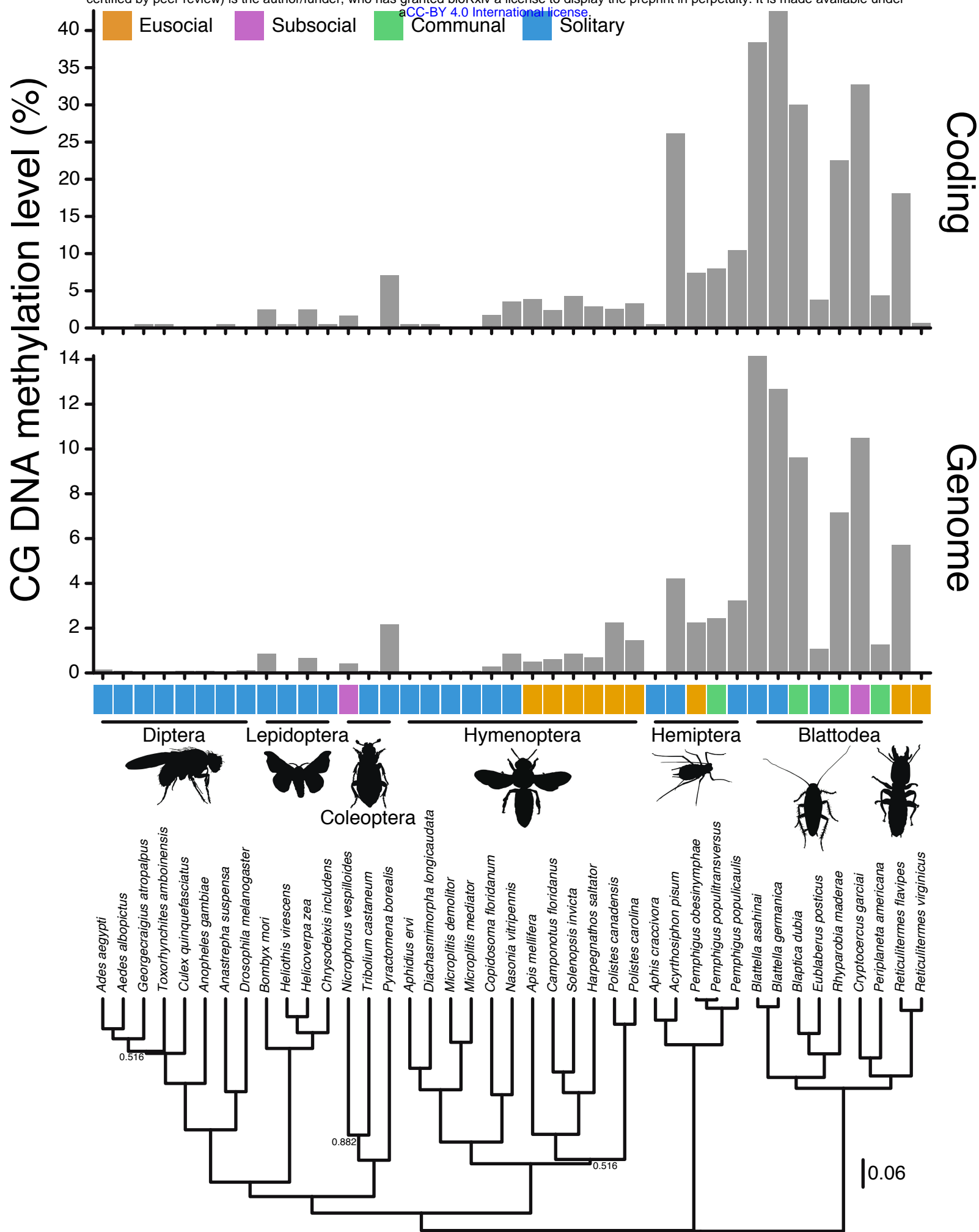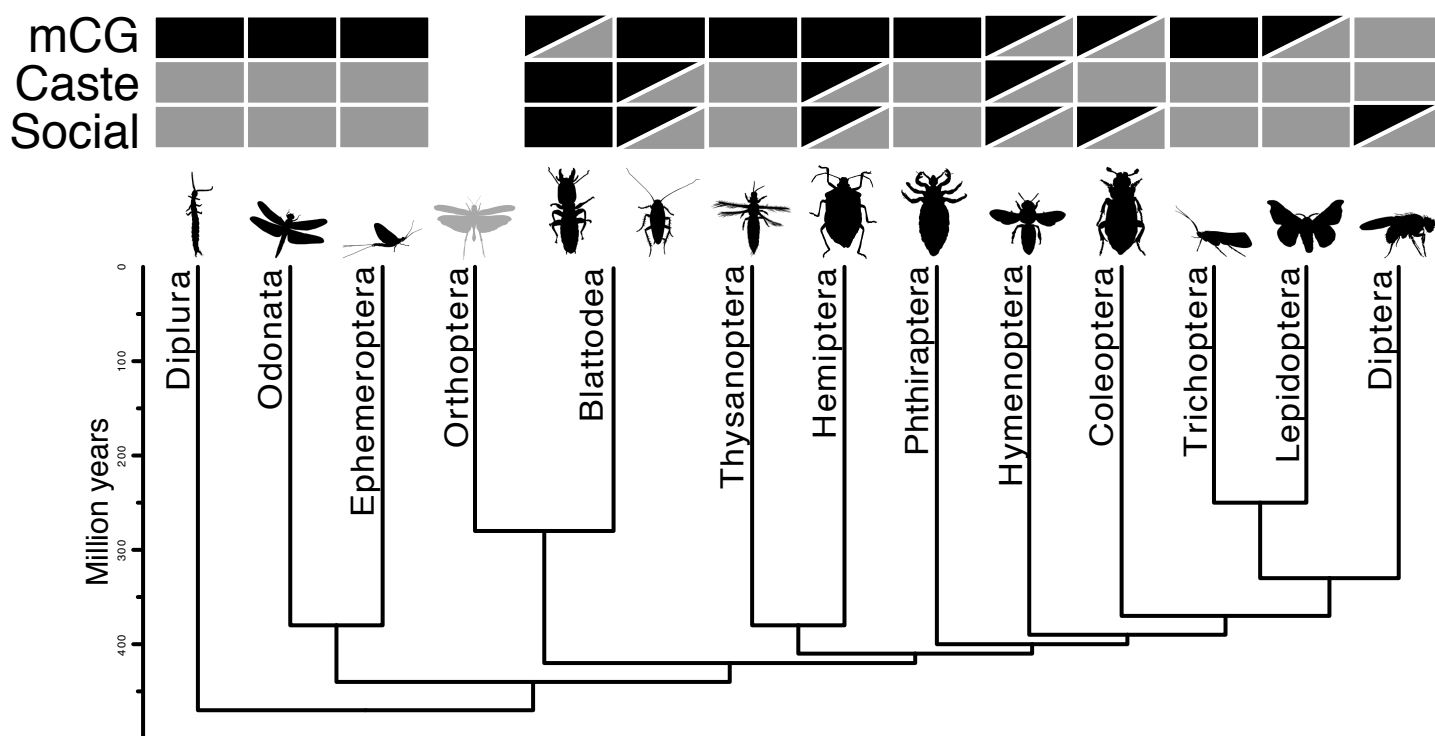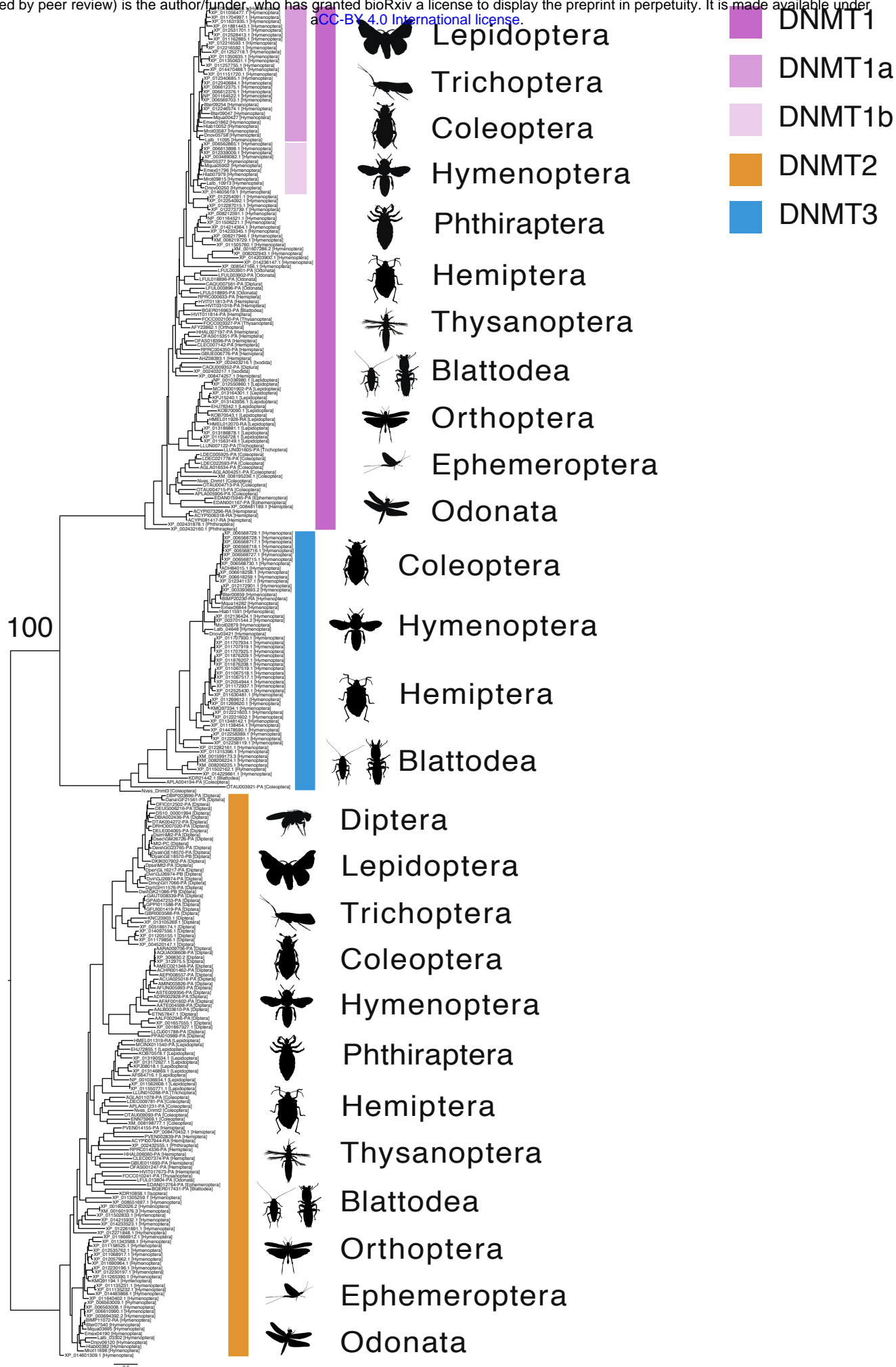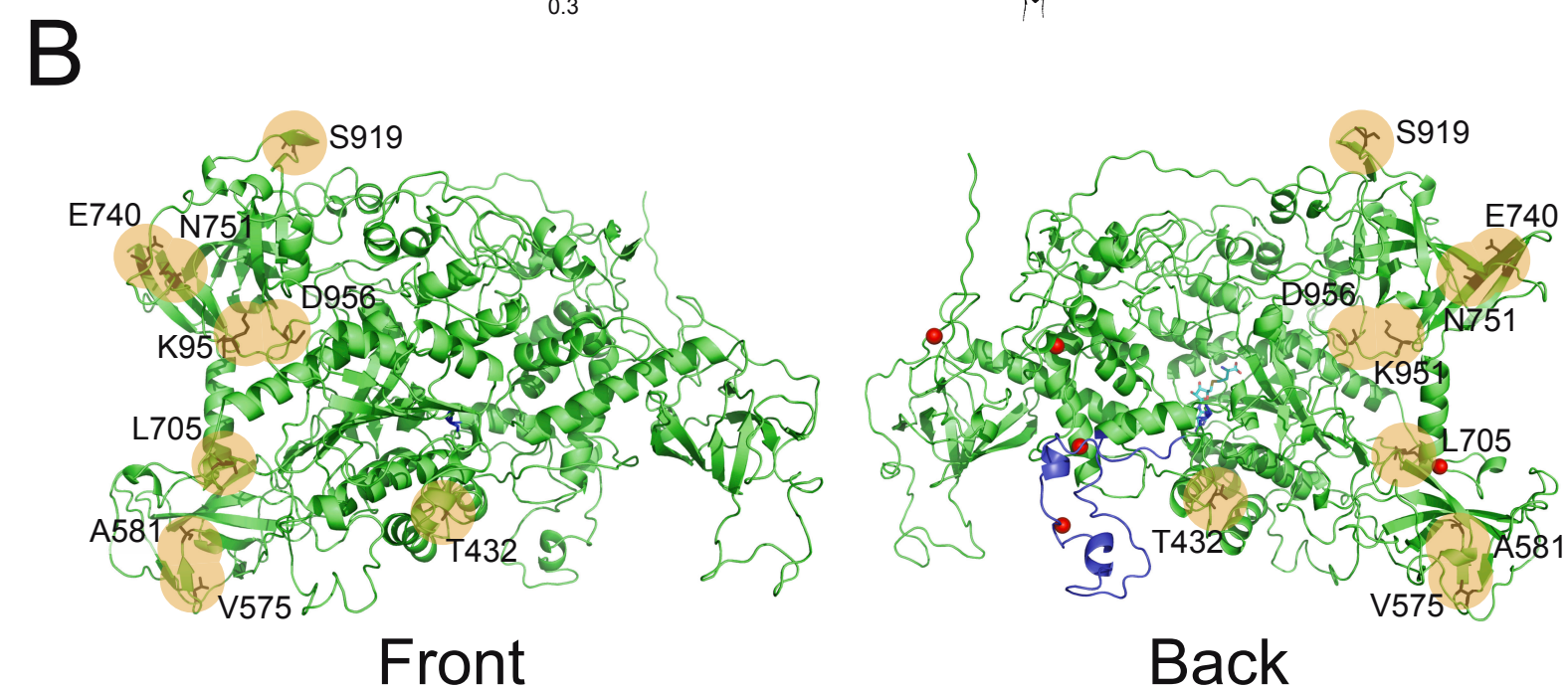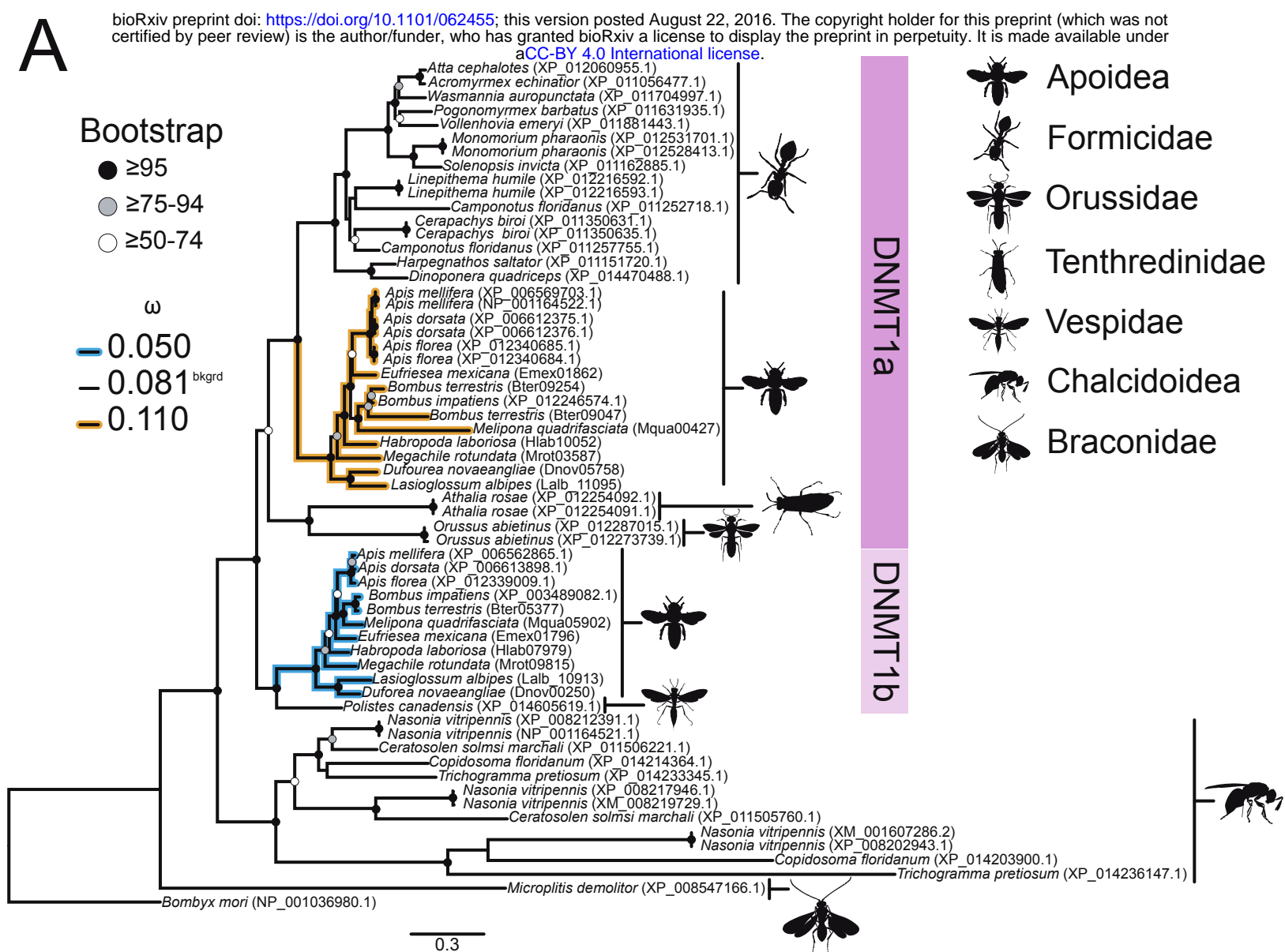
SUPPORTING INFORMATION

Figure S1. $CpG_{O/E}$ **distributions for 123 insect species**. Gaussian mixture modeling was performed in R v2.3.4 to estimate uni- and bimodality. The dashed line represents the mean of each distribution, and the shaded area is the 95% confidence interval around the mean.

Figure S2. **Character state transition matrices from Pagel's test for evolutionary dependence between two binary traits.** Transition rates between states are given for independent and dependent models of evolution. Greyed rates represent a rate of zero (0) between states.

Figure S3. **Genomic levels of DNA methylation are a good predictor of genic levels of DNA methylation.** Linear regression between genomic and genic levels of DNA methylation for 22 species of insects with sequenced genomes.

Figure S4. **The duplication of DNMT1 in Hymenoptera was shared by ants and bees.** (A) The hypothesized relationships of DNMT1 in Hymenoptera. The maximum likelihood tree was generated using an amino acid alignment and RAxML with the PROTGAMMAWAG model of protein evolution. Selected values at nodes represent bootstrap support from 1000 replicates. (B) Superfamily and family relationships suggest the duplication of DNMT1 in Hymenoptera was shared by ants and bees, and reciprocal losses of paralogs (DNMT1a and DNMT1b) occurred in Formicidae and Vespidae.

Table S1. Species and accompanying behavior, genomic, DNA methylation and sequencing information used in this study. Rows shaded in grey represent the 18 species where $CpG_{O/E}$ correctly predicted presence/absence of DNA methylation based on comparisons to WGBS.

Table S2. Species and DNA methyltransferase (DNMT) accession id's used in Figure 3.

Table S3. A list of proteins used to estimate multilocus coalescent trees and the best BLAST hit to *Apis mellifera* or *Drosophila melanogaster*.

Table S4. Hypotheses tested using Phylogenetic Analysis using Maximum Likelihood (PAML) for DNMT1a and DNMT1b in Apoidea.

REFERENCES

Attardo GM, Benoit JB, Michalkova V, Patrick KR, Krause TB, Aksoy S. 2014. The homeodomain protein ladybird late regulates synthesis of milk proteins during pregnancy in the tsetse fly (*Glossina morsitans*). PLoS Negl Trop Dis. 8:e2645.

Blackman RL, Eastop VF. 2000. Aphids on the world's crops: an identification and information guide. John Wiley & Sons Inc.

Dressler RL. 1985. Euglossine bees (Hymenoptera: Apidae) of the Tambopata reserved zone, Madre de Dios, Perú. Rev Peruana Entomol. 27:75-79.

Kocher SD, Li C, Yang W, Tan H, Yi SV, Yang X, Hoekstra HE, Zhang G, Pierce NE, Yu DW. 2013. The draft genome of a socially polymorphic halictid bee, *Lasioglossum albipes*. Genome Biol. 14:R142.

Monnin T, Peeters C. 1998. Monogyny and regulation of worker mating in the queenless ant *Dinoponera quadriceps*. Anim Behav. 55:299–306.

Scott MP. 1998. The ecology and behavior of burying. Annu. Rev. Entomol. 43:595–618.

Strand MR, Grbic M. 1997. The development and evolution of polyembryonic insects. Curr Top Dev Biol. 35:121–159.

Tsuji K, Yamauchi K. 1995. Production of females by parthenogenesis in the ant, *Cerapachys biroi*. Insectes Soc. 42:333–336.

Uzsák A, Schal C. 2013. Sensory cues involved in social facilitation of reproduction in *Blattella germanica* females. PLoS One. 8:e55678.

Blattella germanica
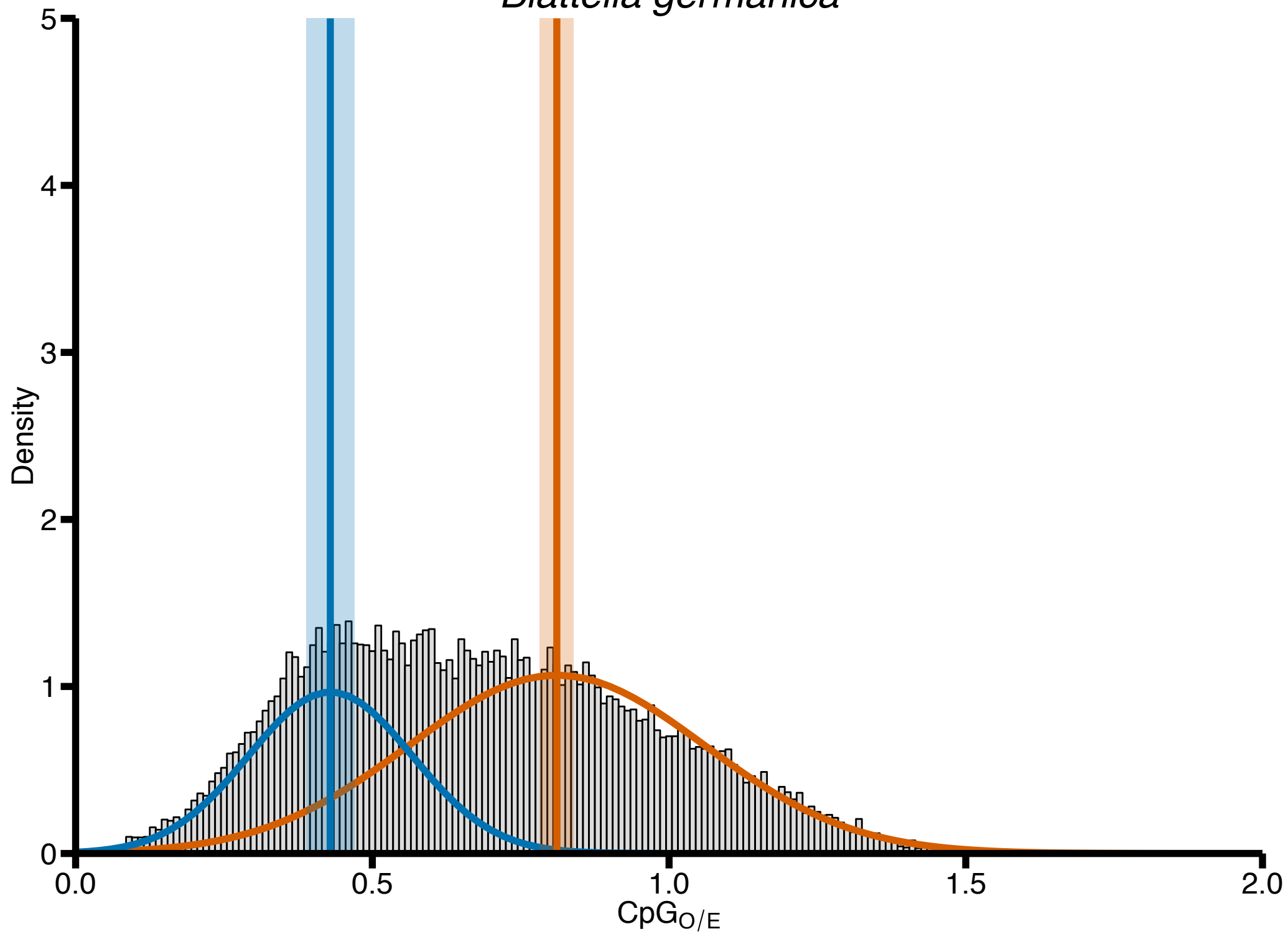
Agrilus planipennis
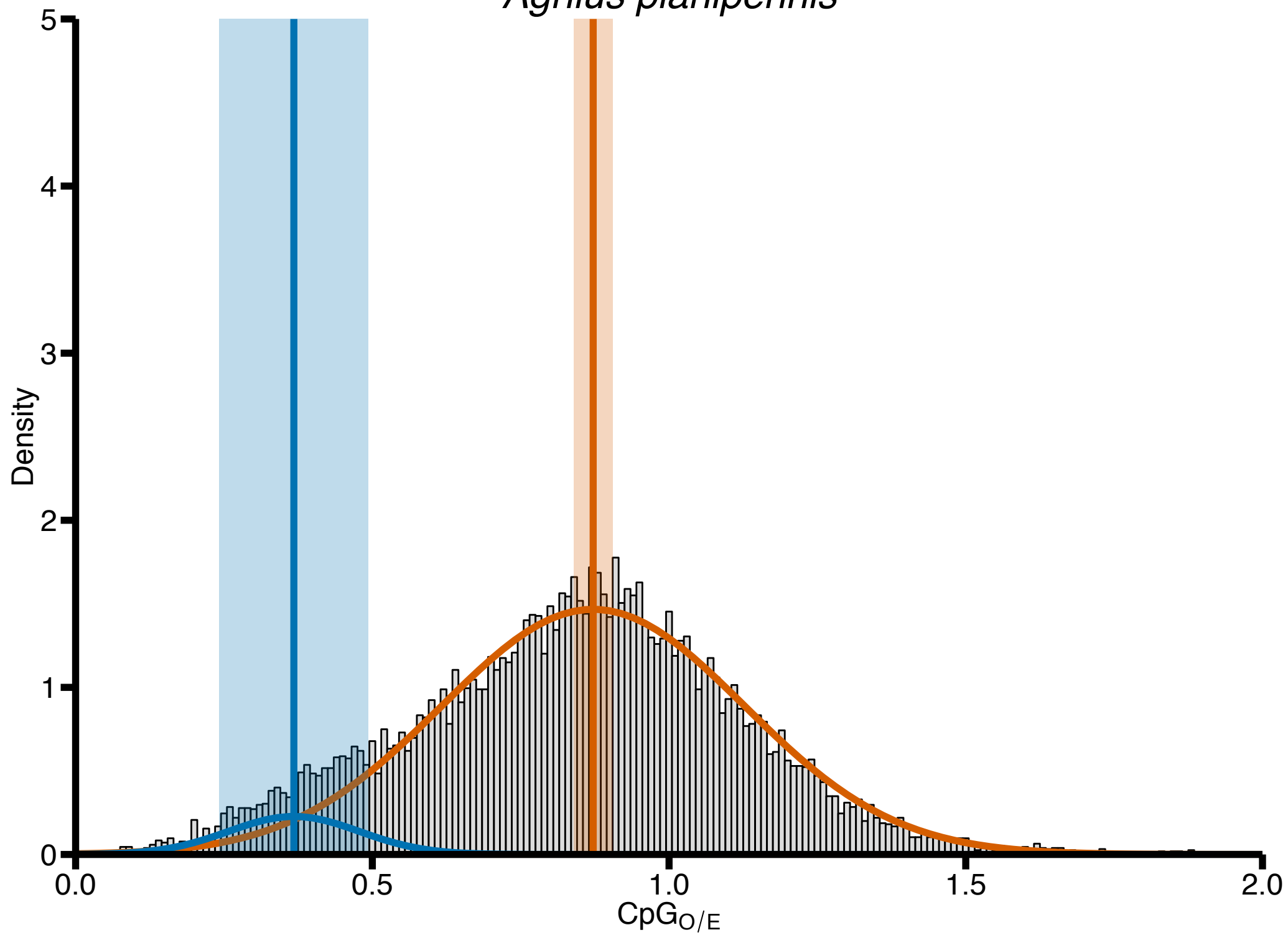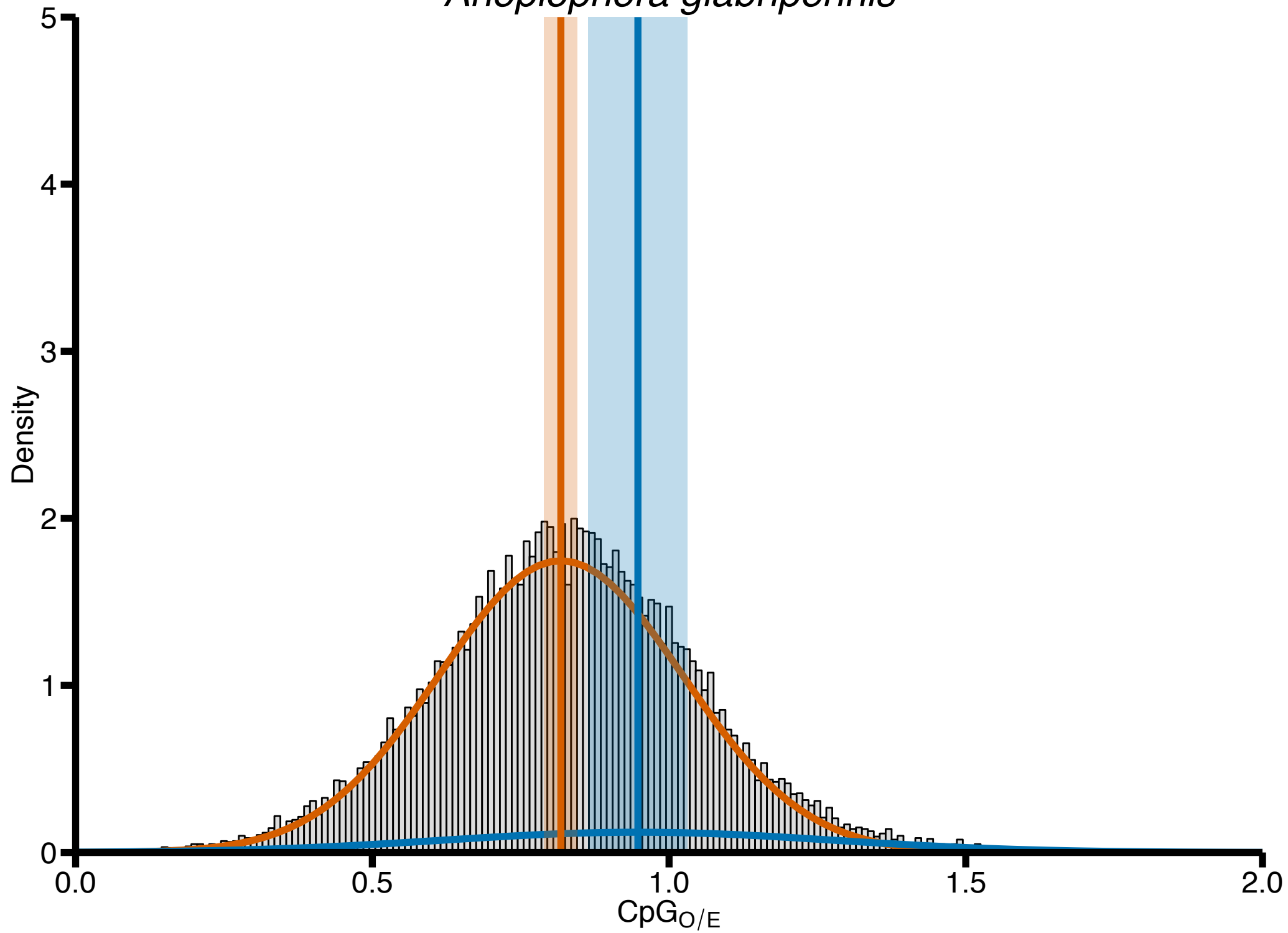
Dendroctonus ponderosae
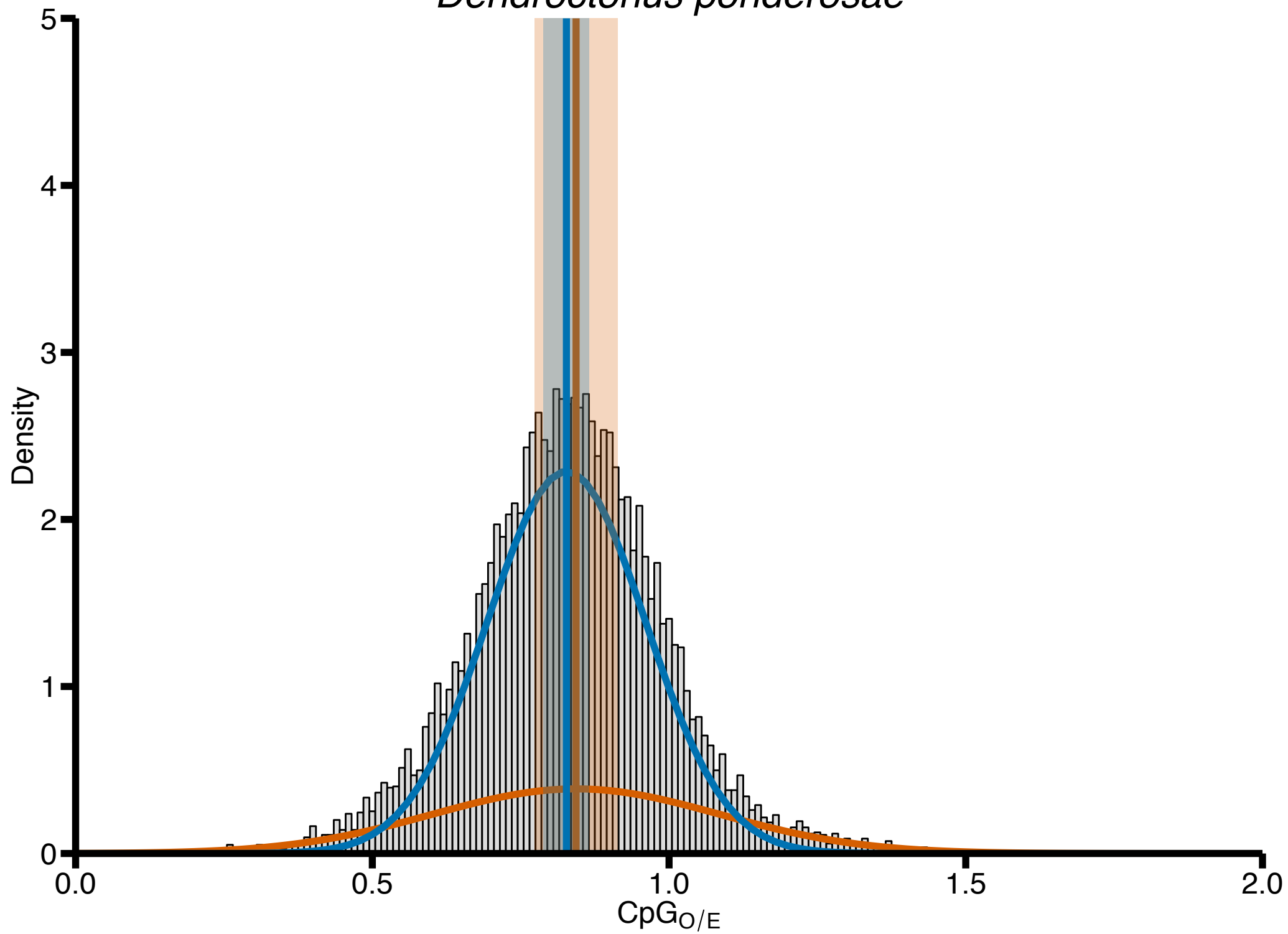
Leptinotarsa decemlineata
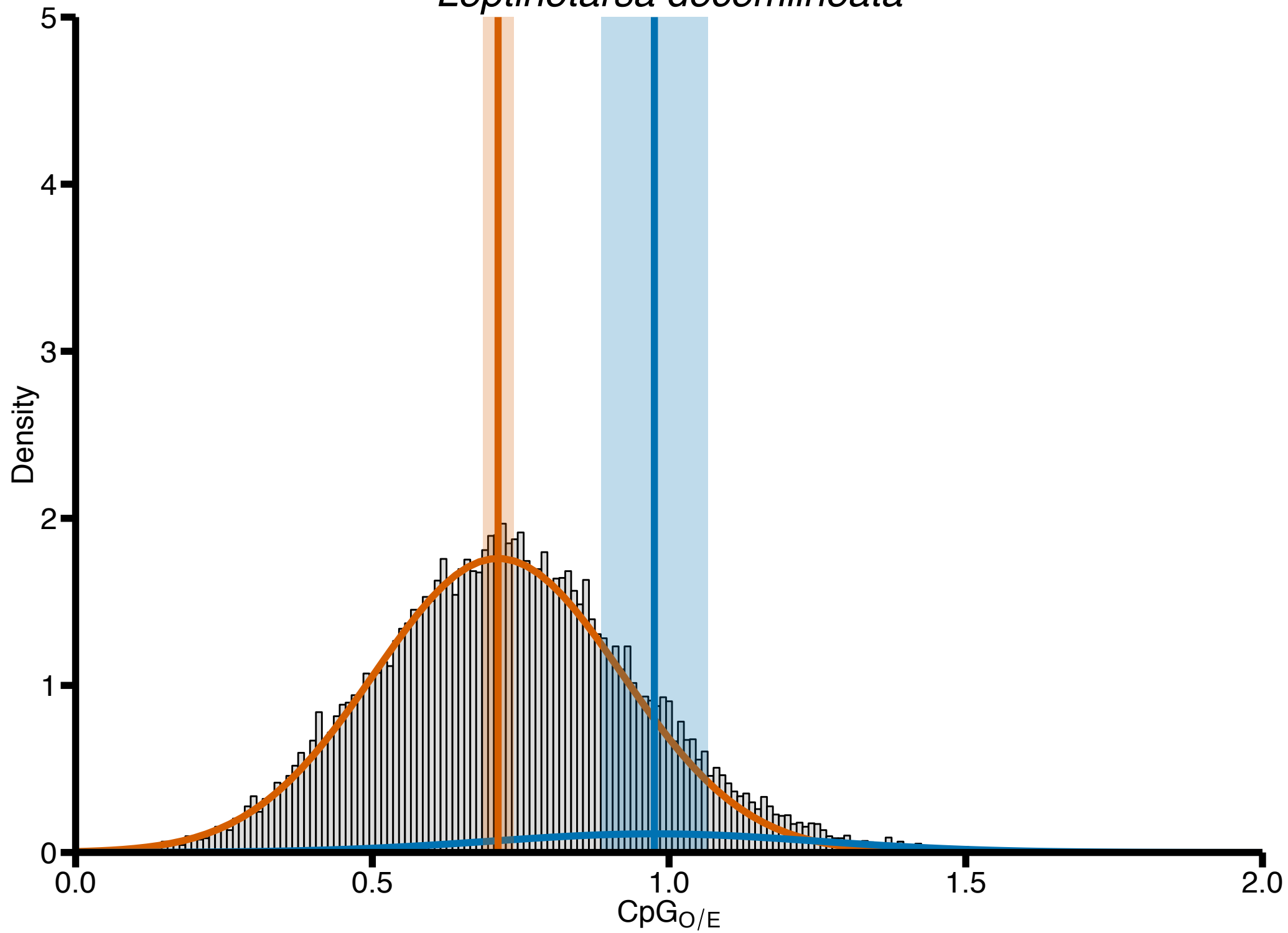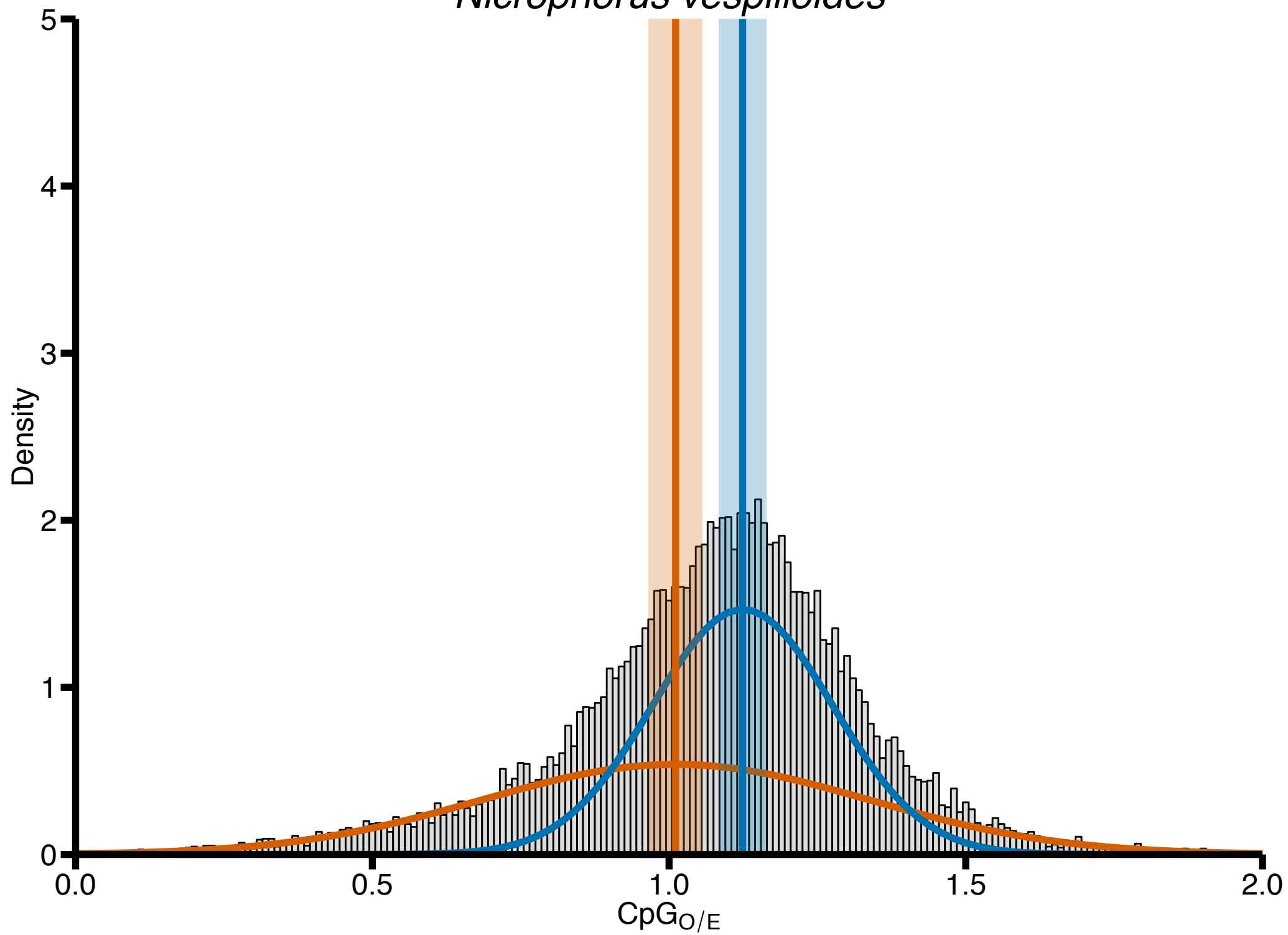
Onthophagus taurus
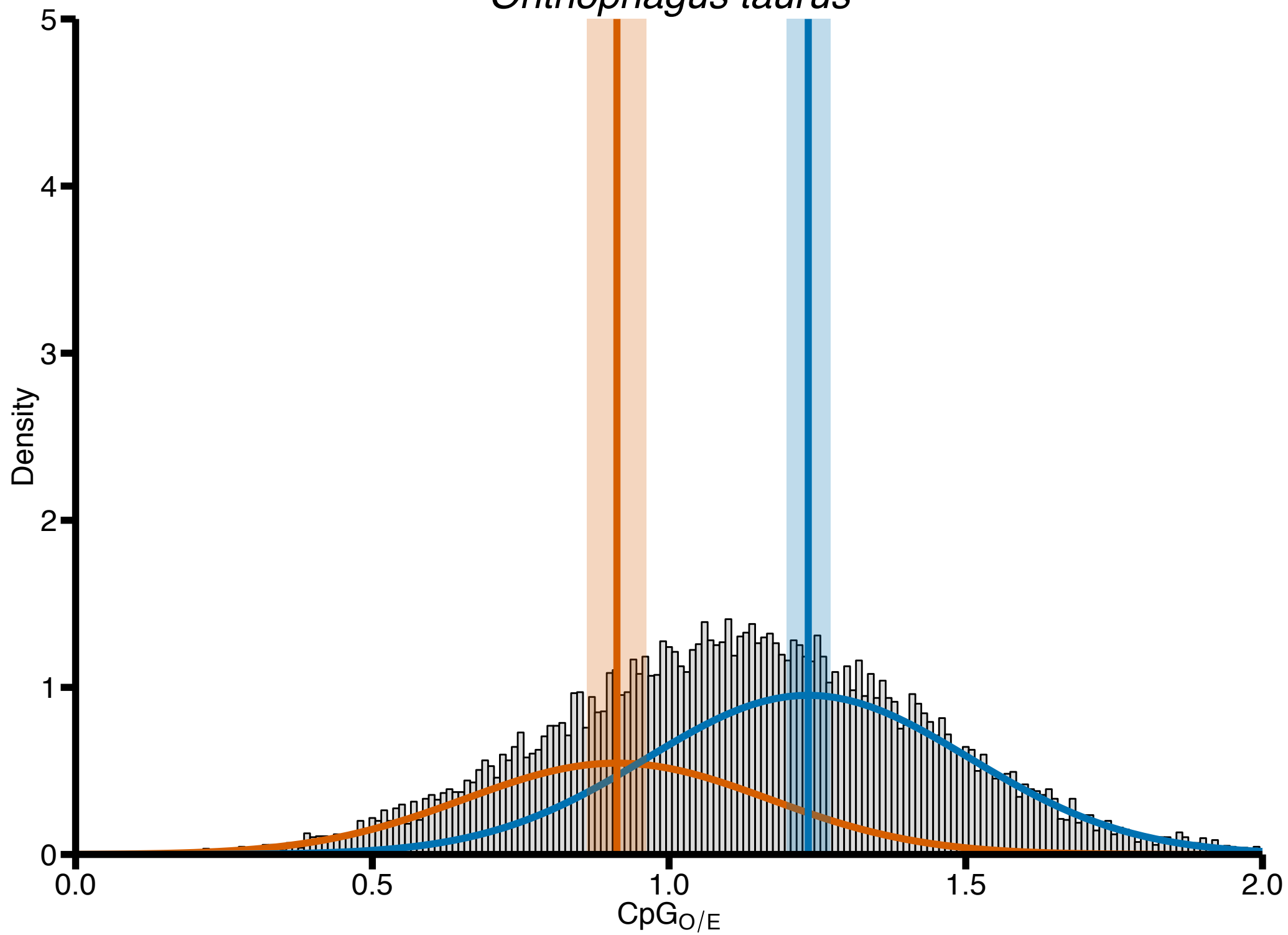
Tribolium castaneum
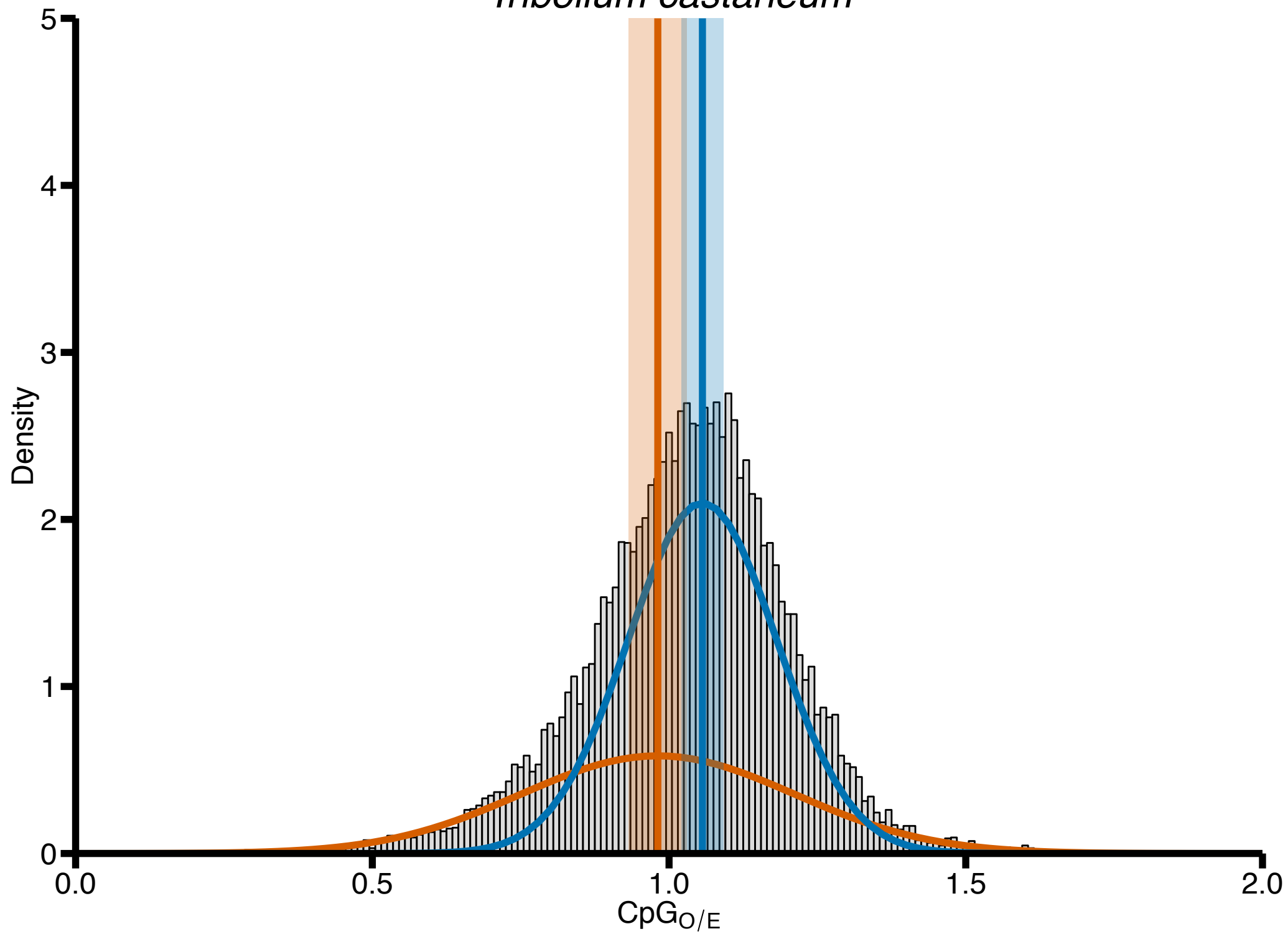
Daphnia pulex
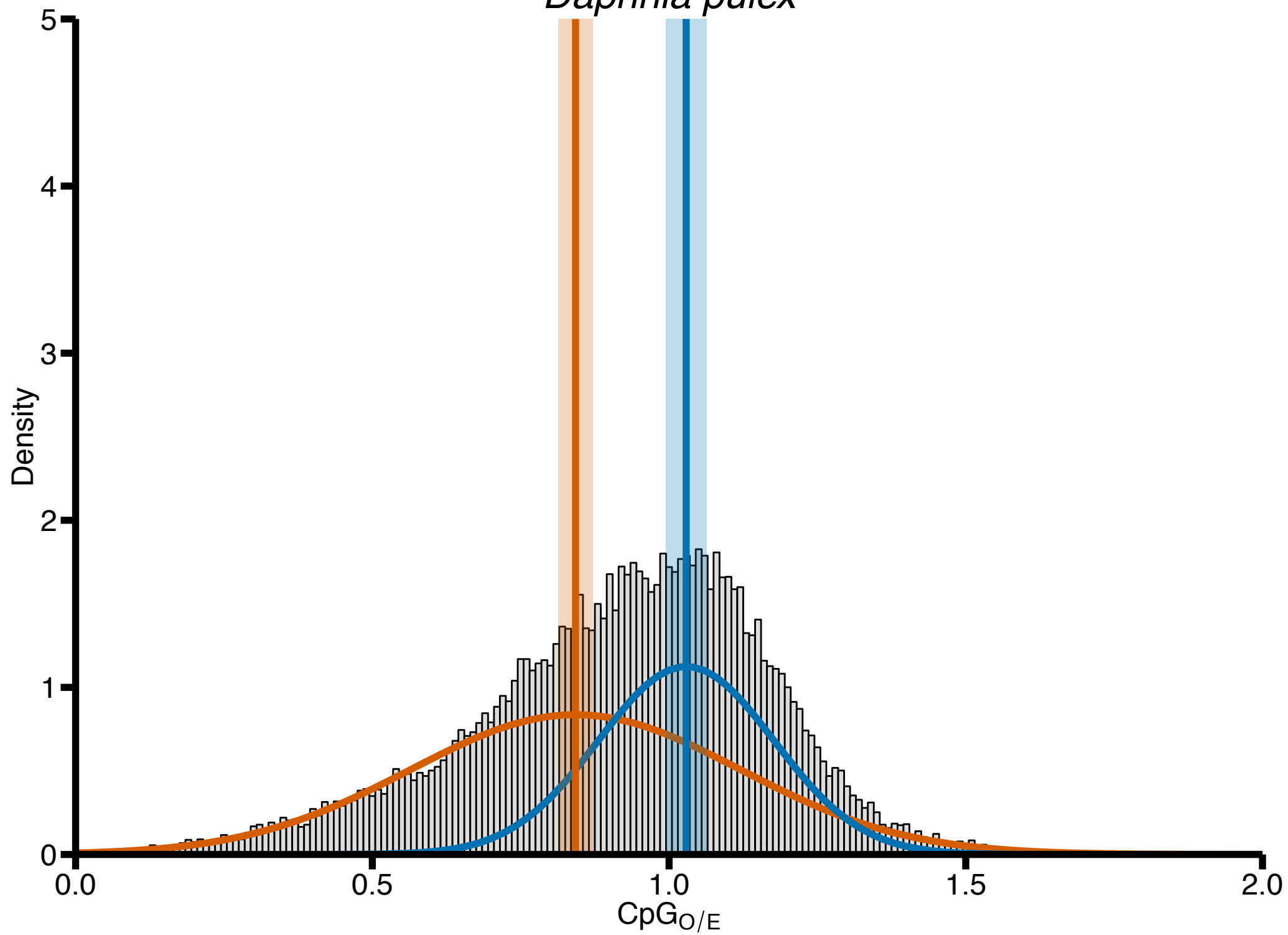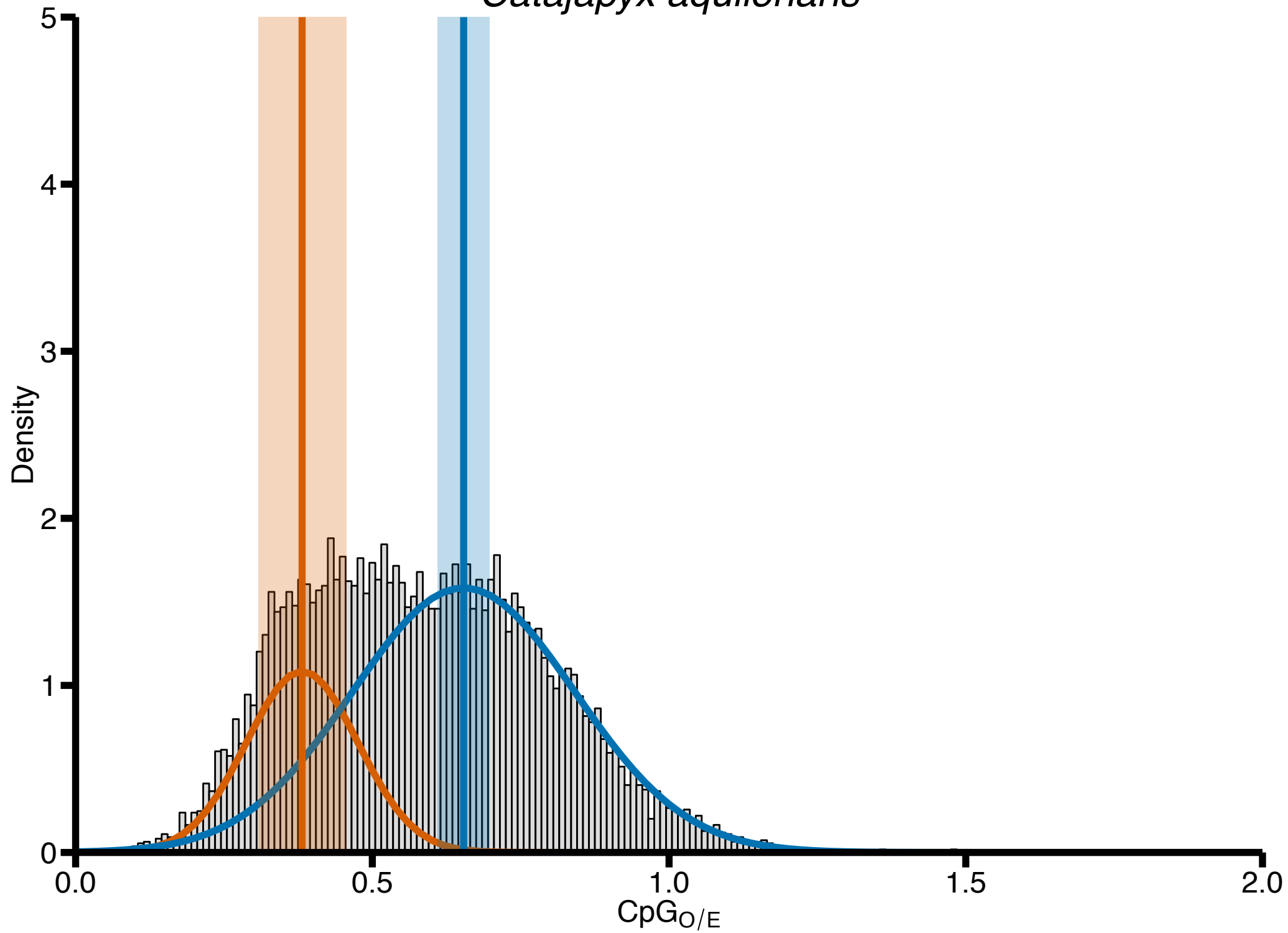
Aedes aegypti
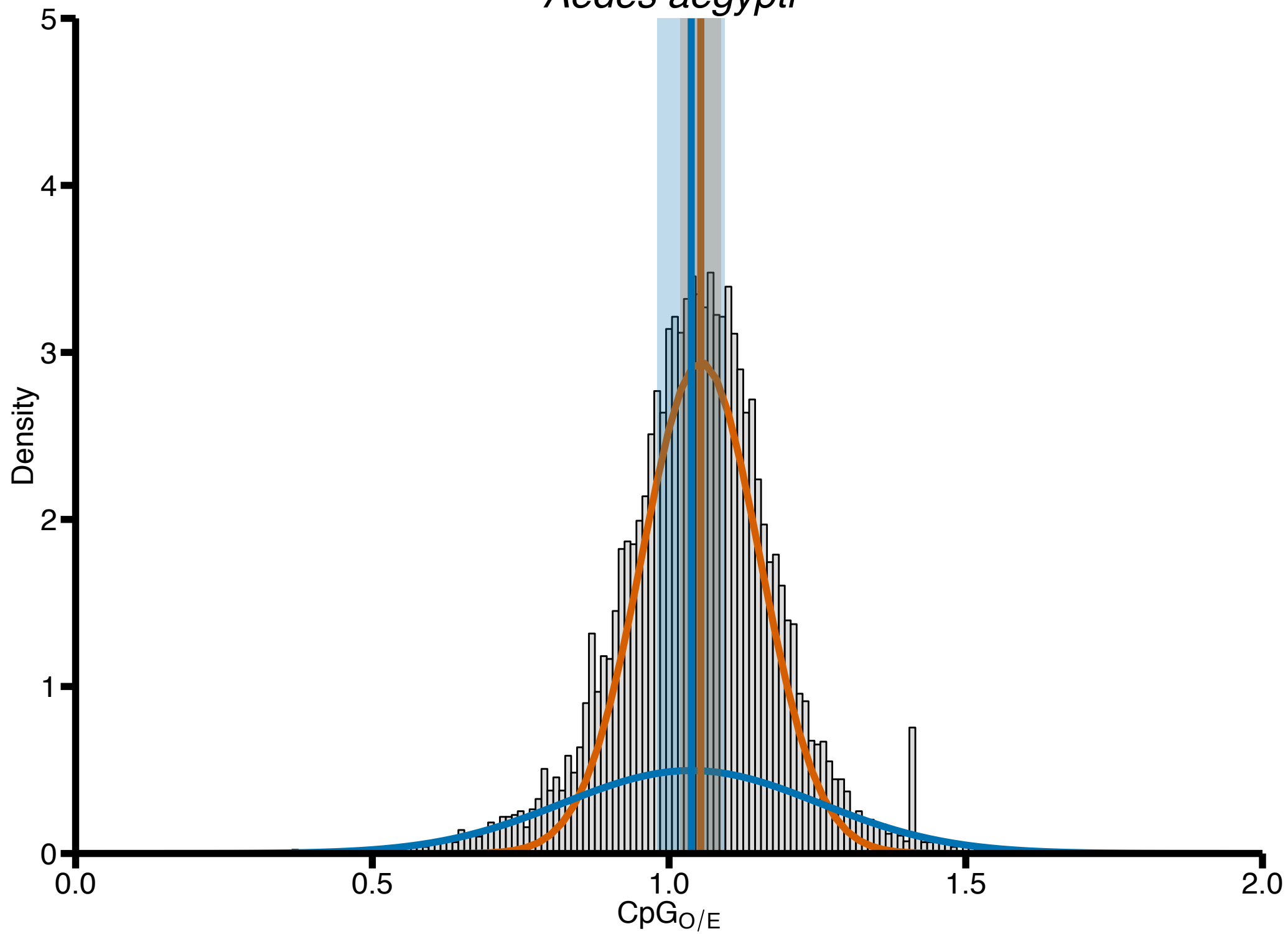
Aedes albopictus
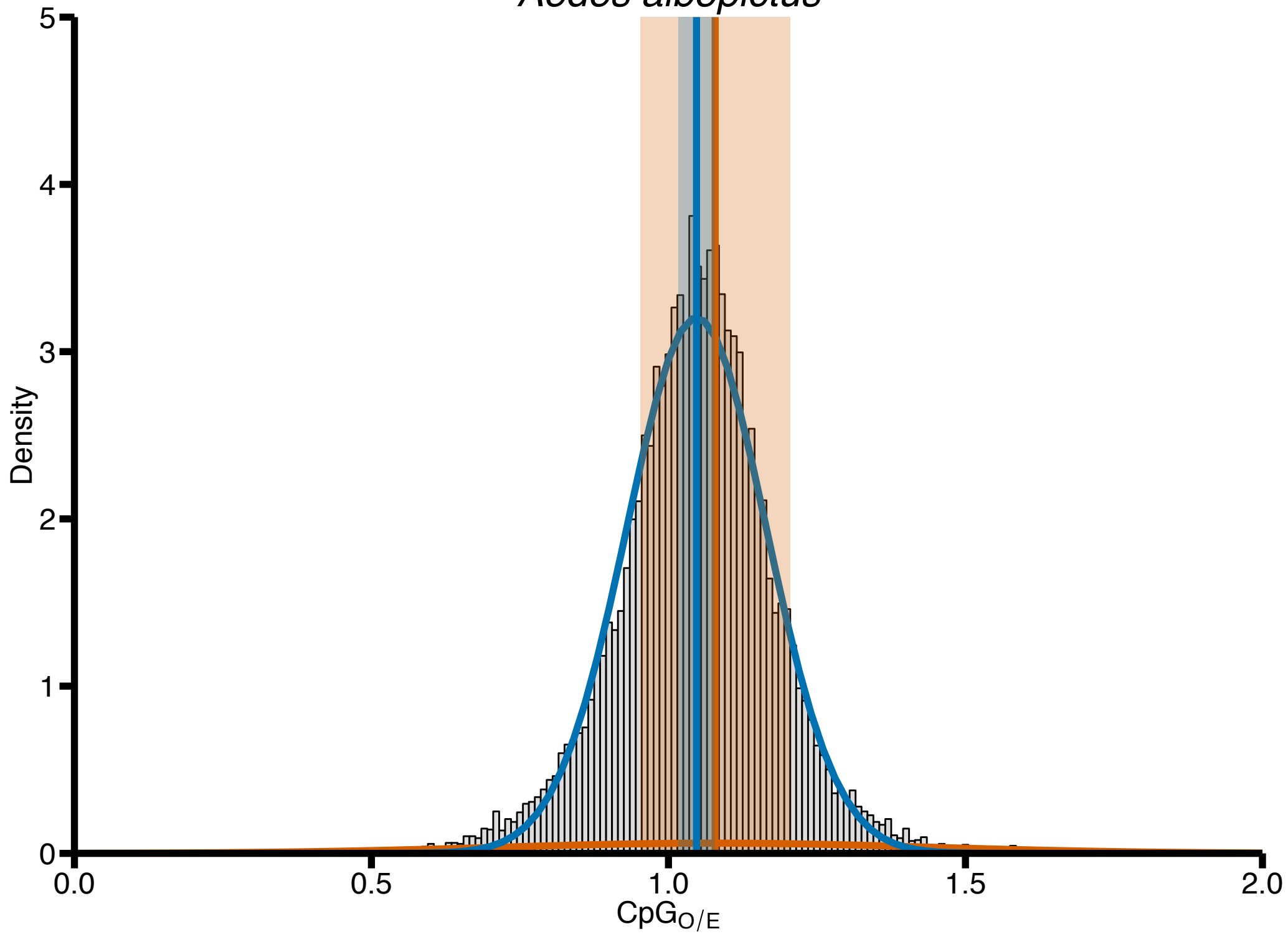
Anopheles albimanus
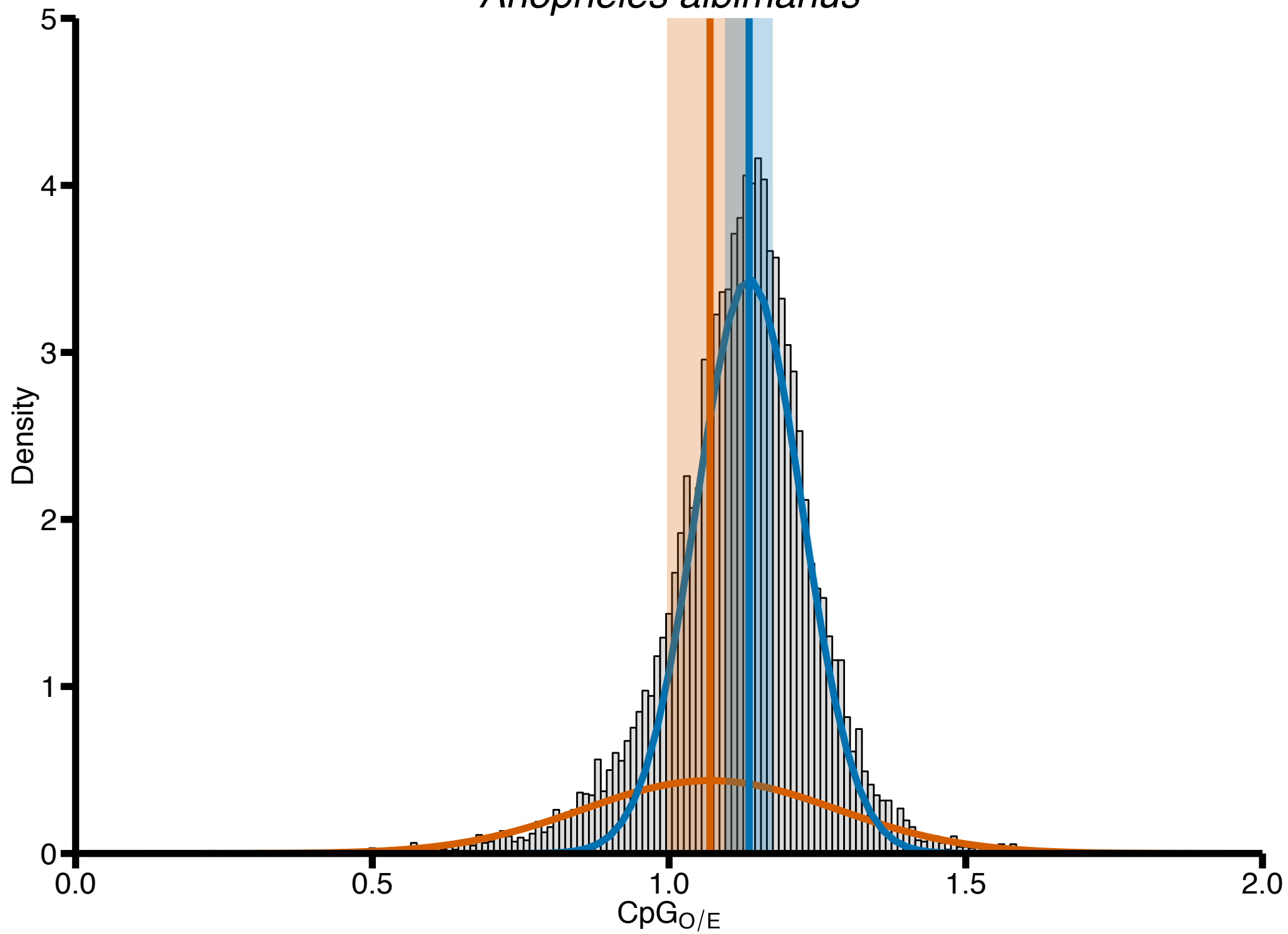
Anopheles arabiensis
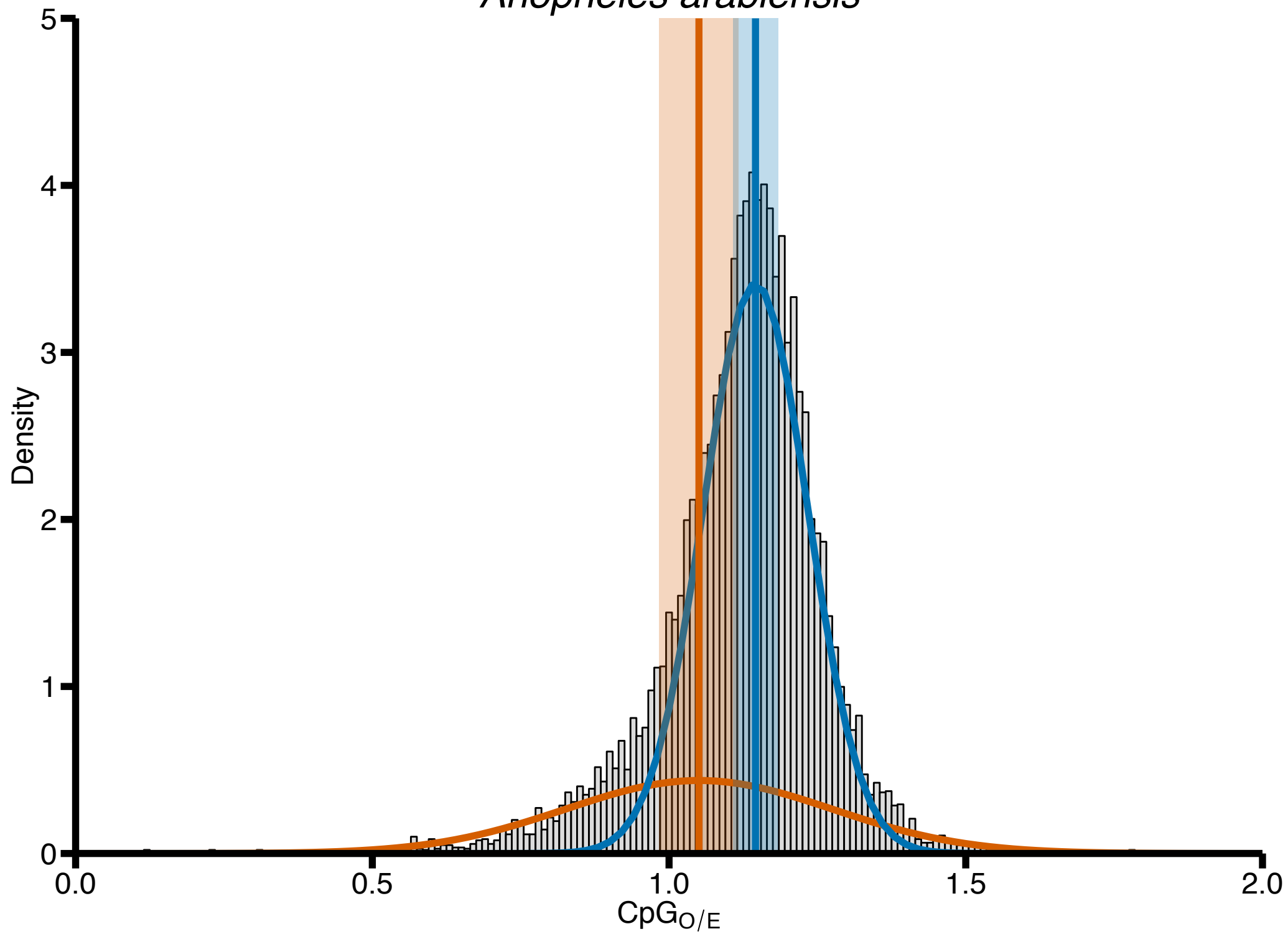
Anopheles atroparvus

Density

CpG$_{O/E}$
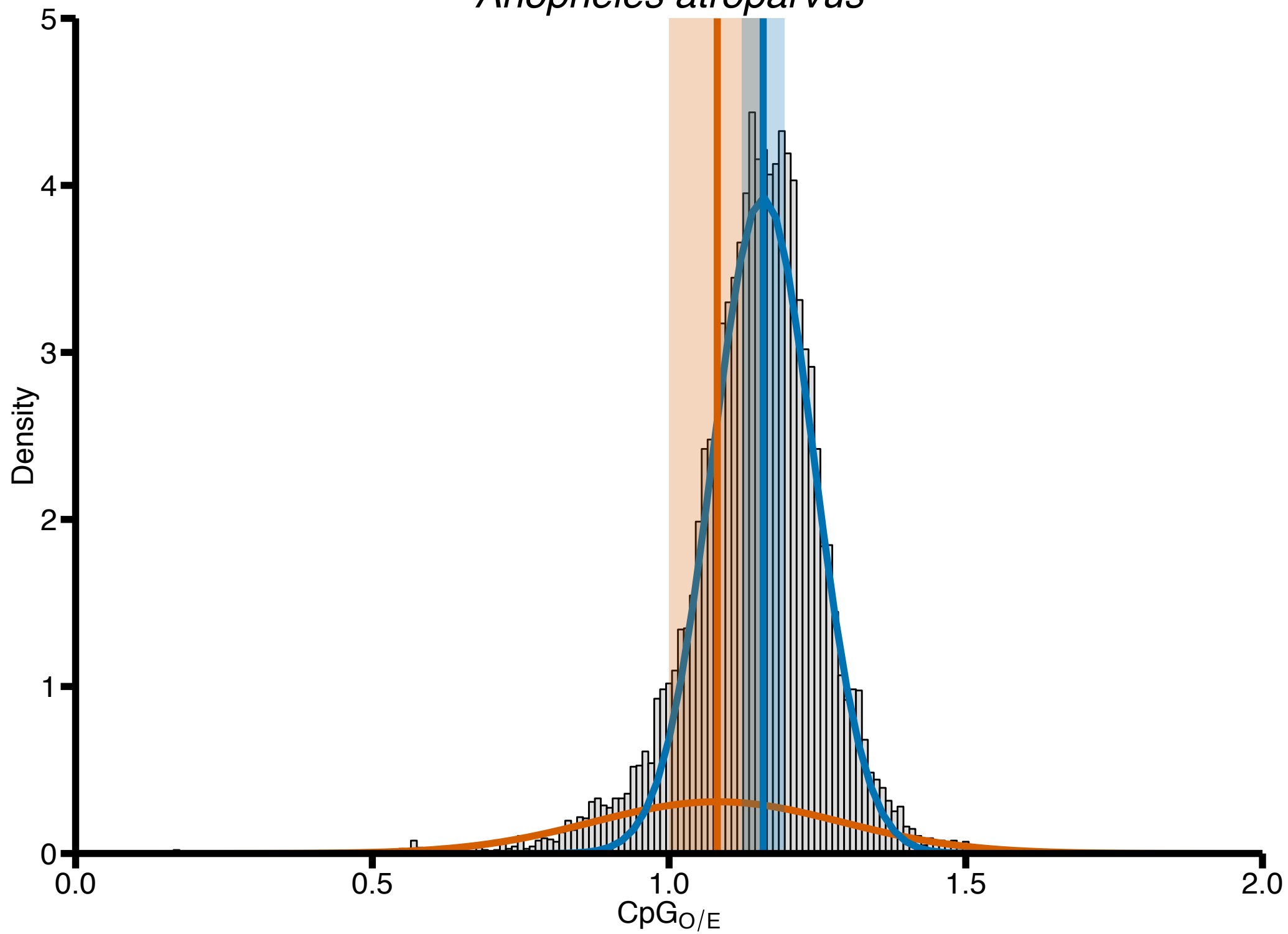
Anopheles christyi
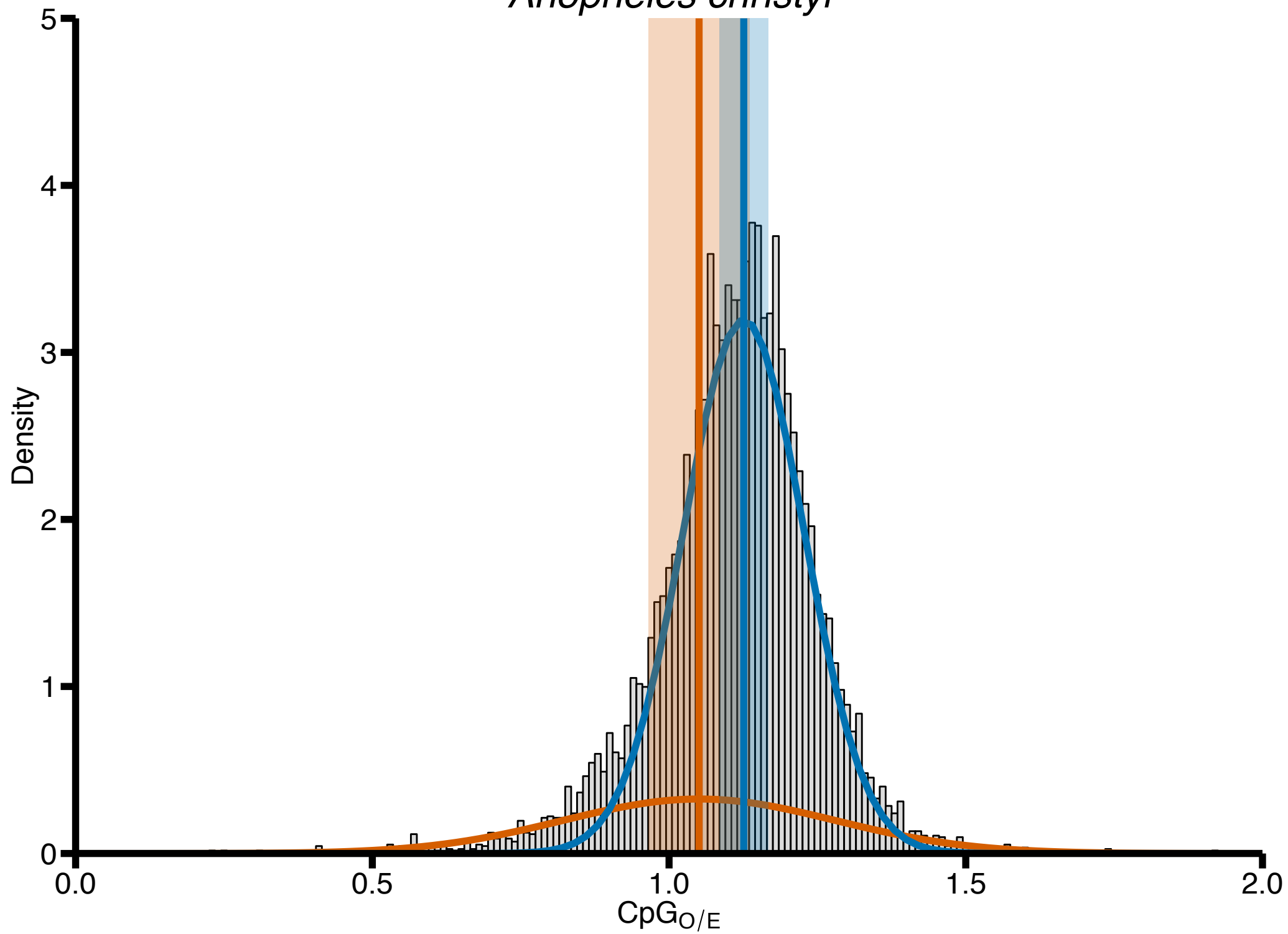
Anopheles culicifacies
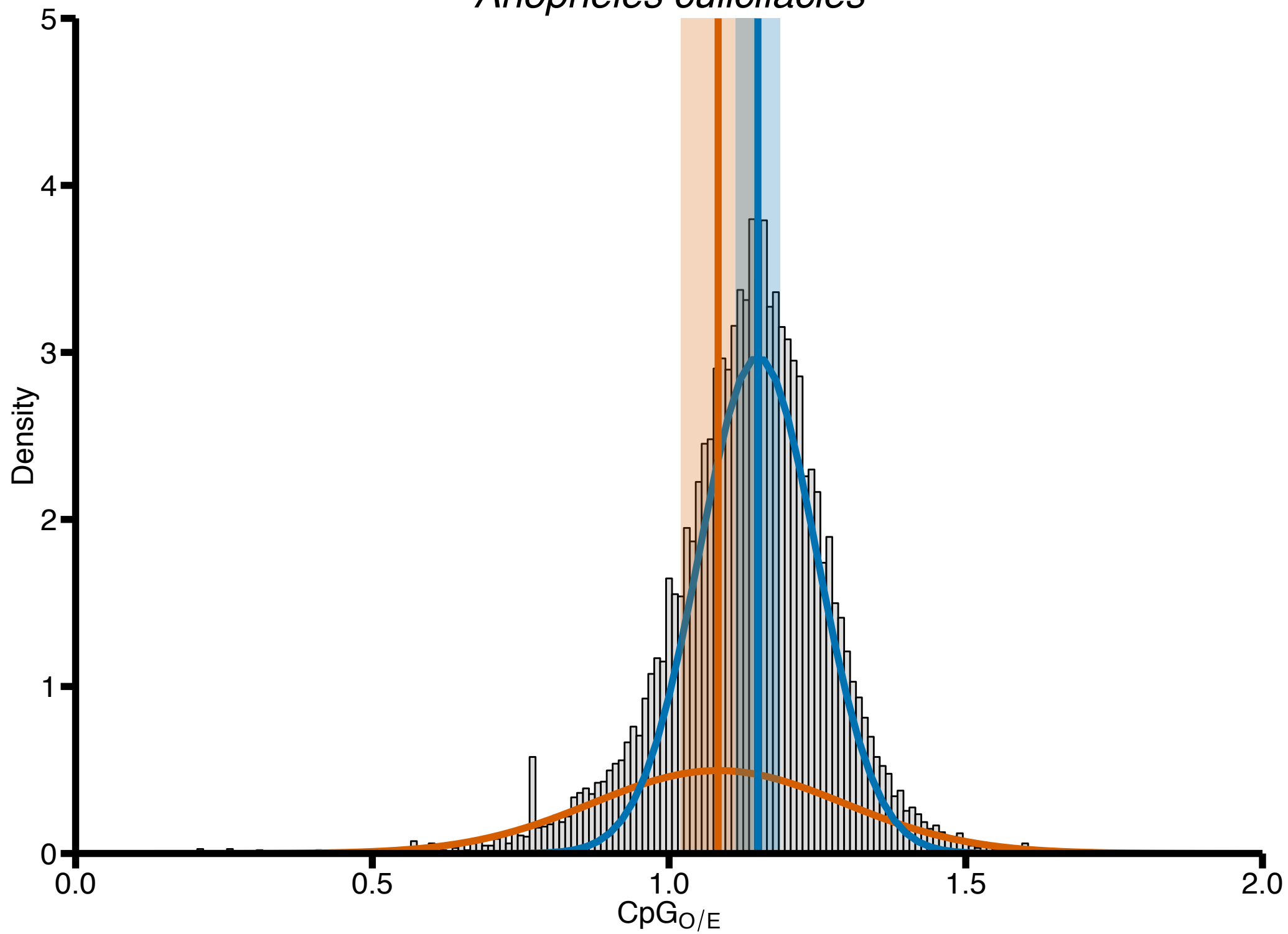
Anopheles darlingi
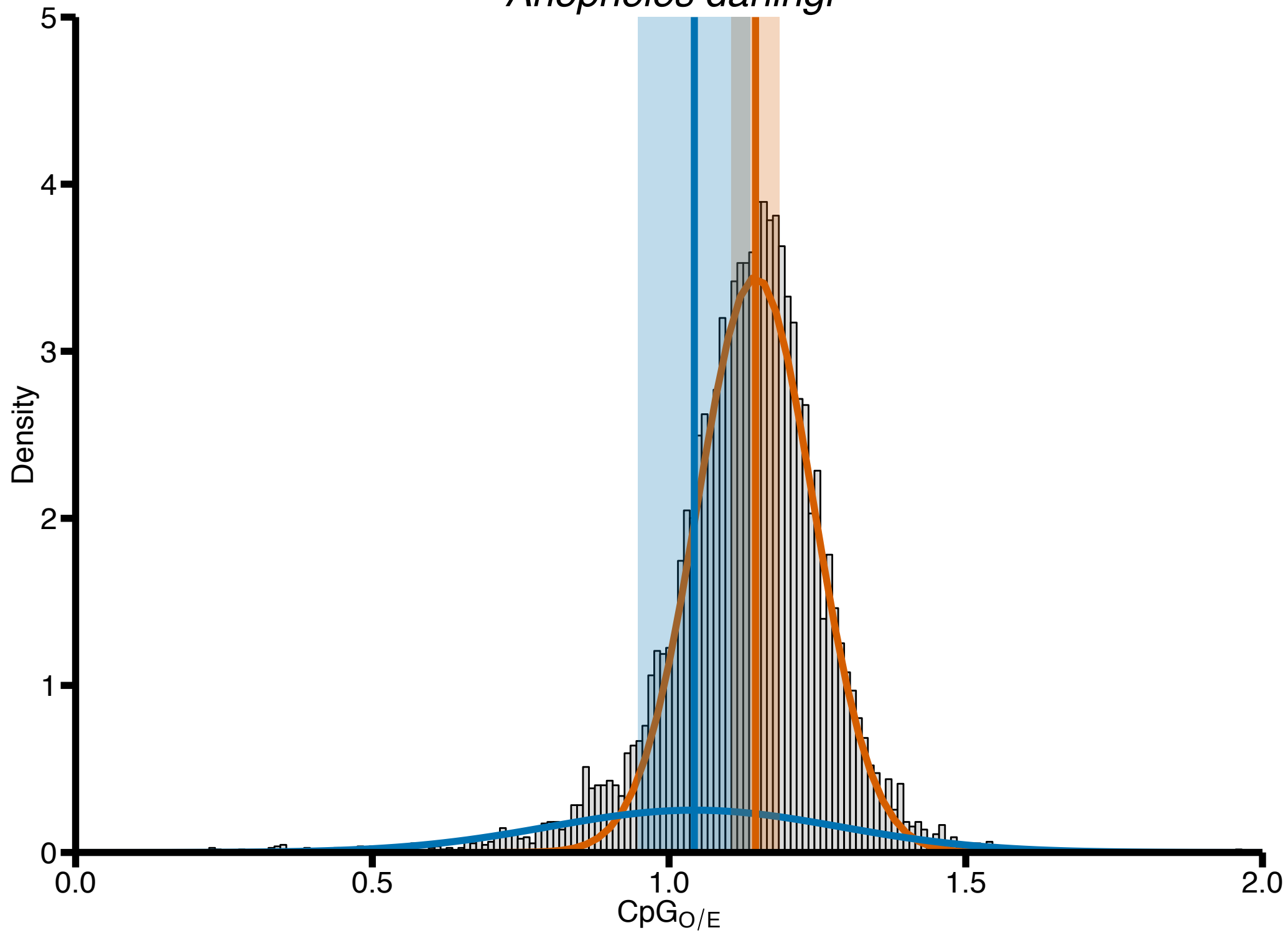
Anopheles dirus
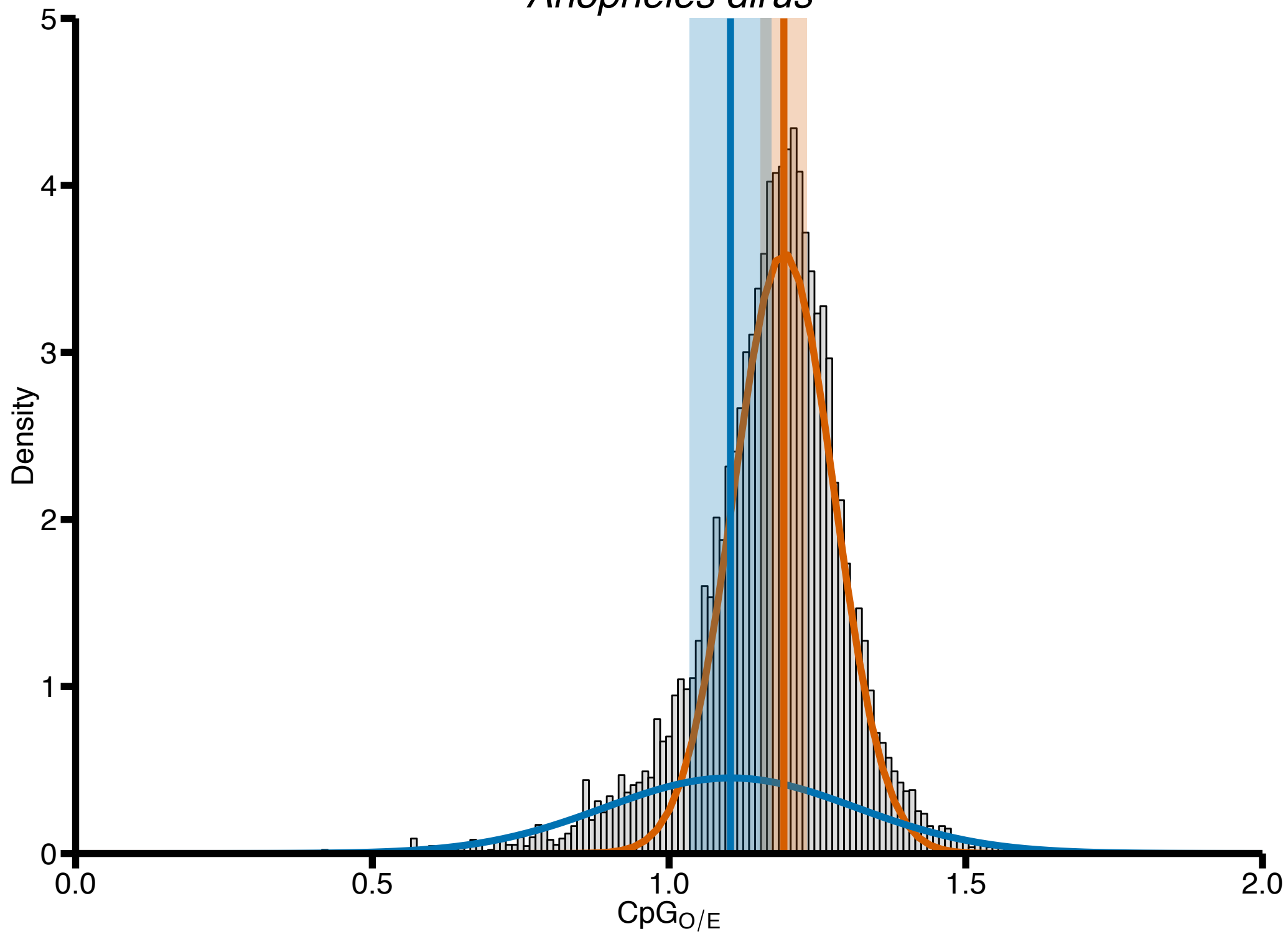
Anopheles epiroticus
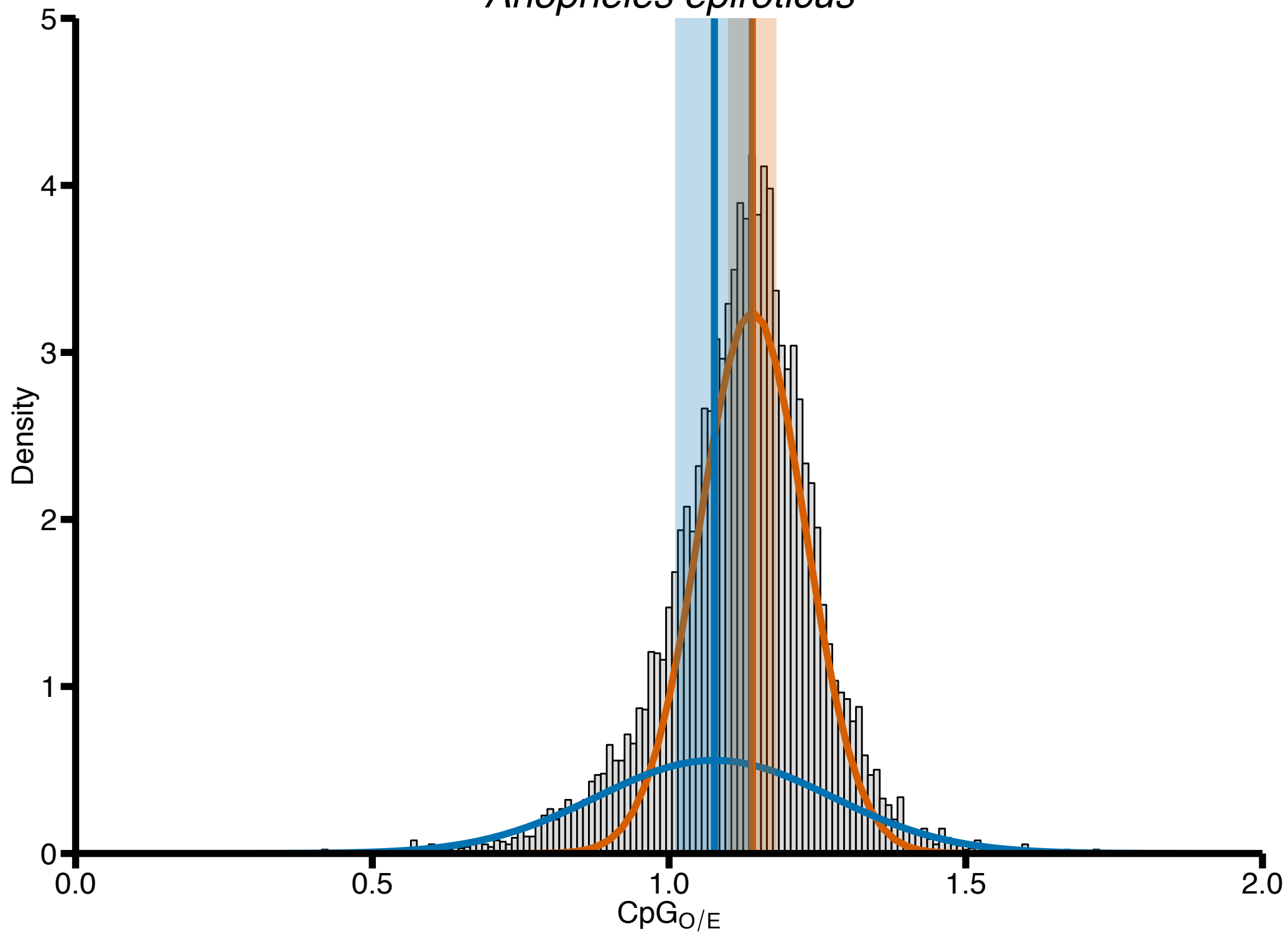
Anopheles farauti
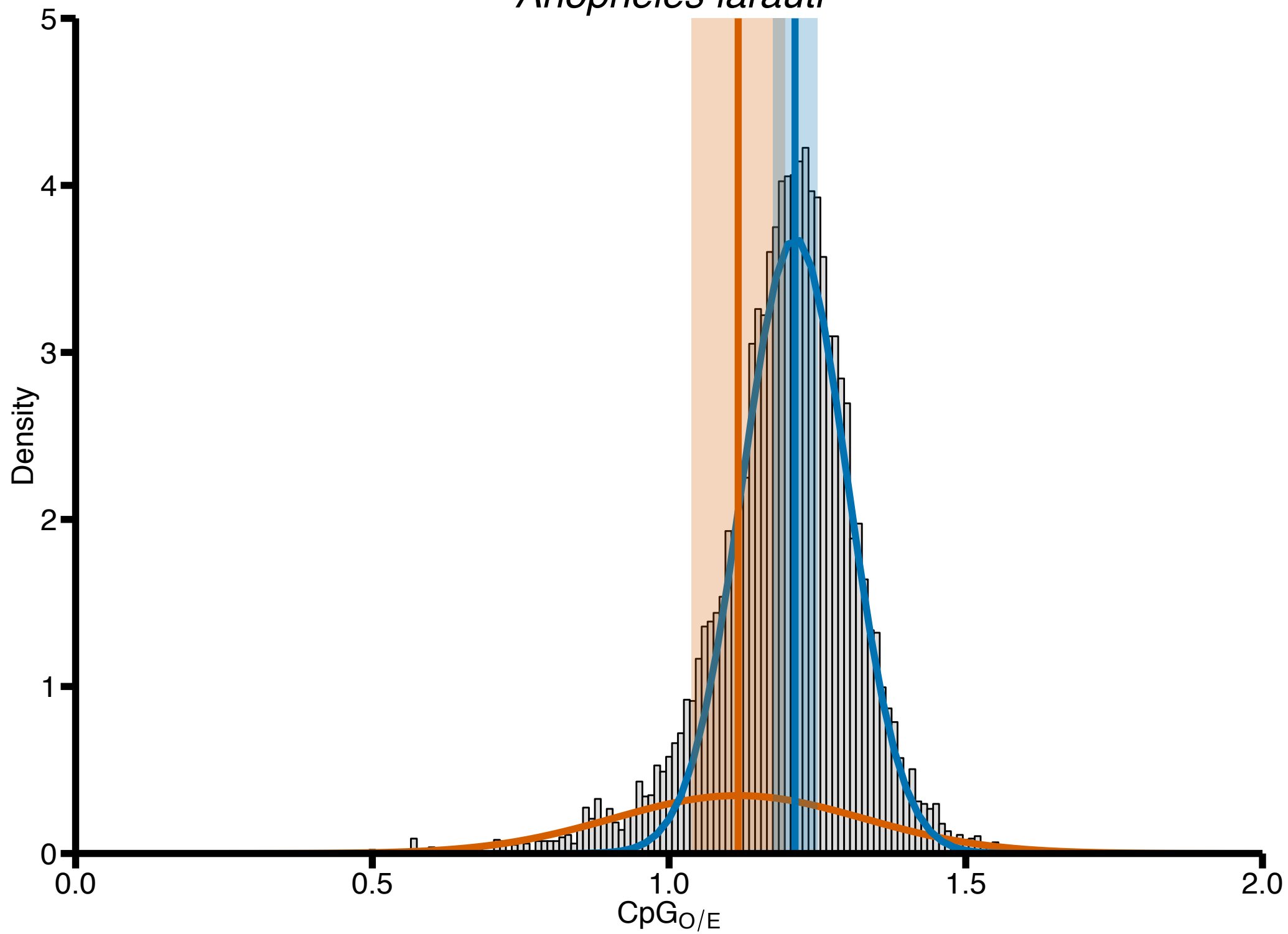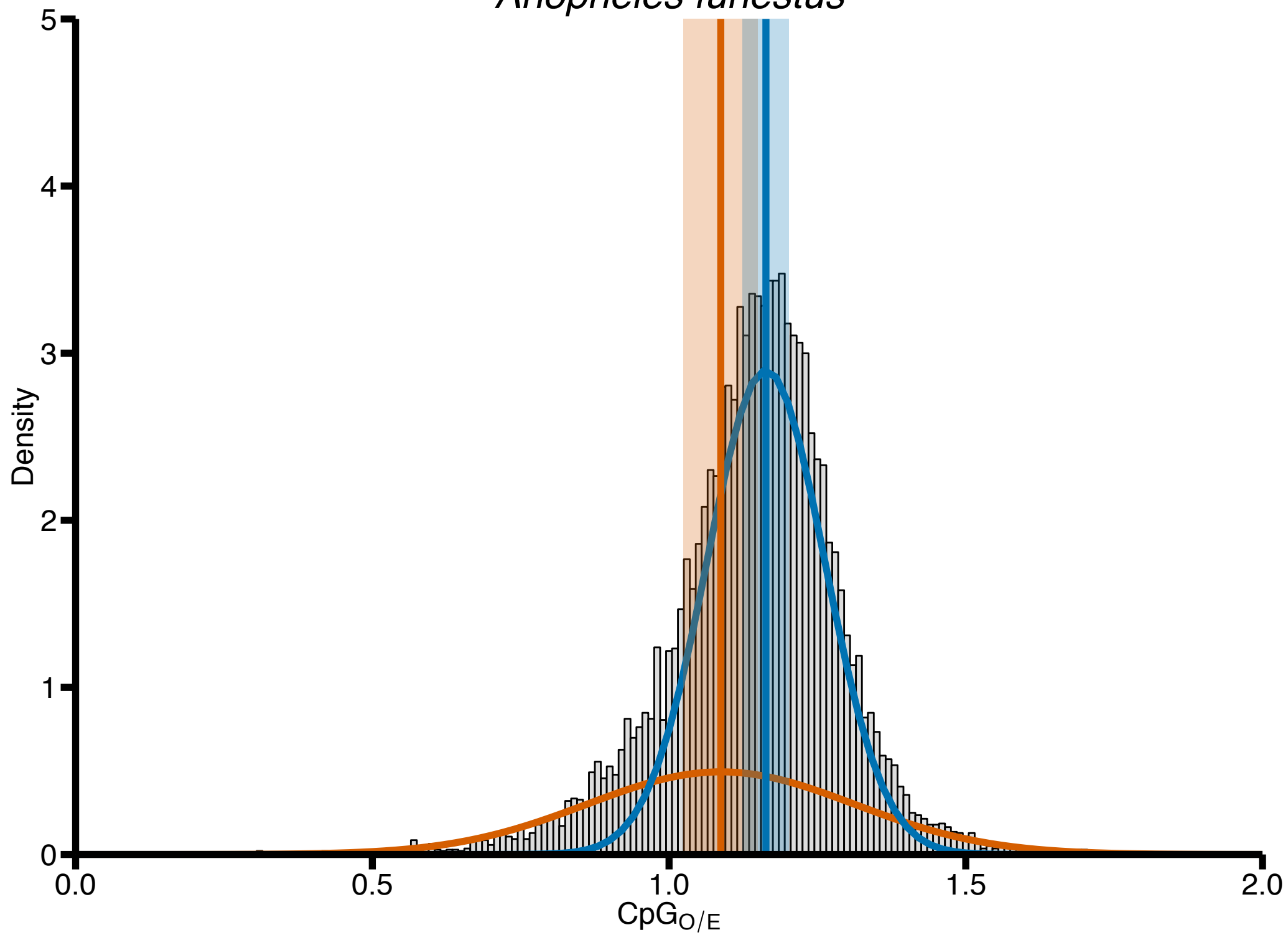
Anopheles gambiae
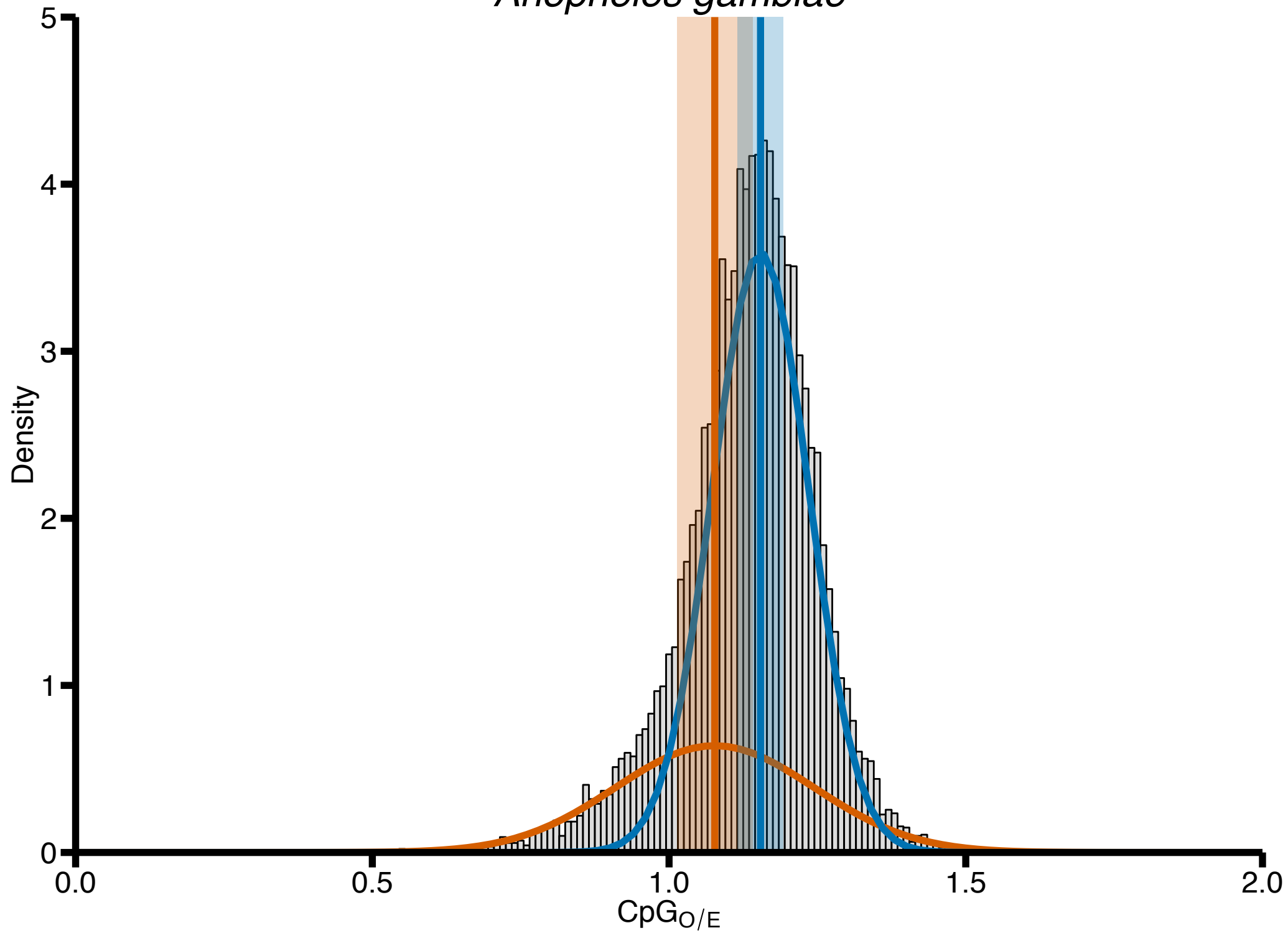
Anopheles maculatus
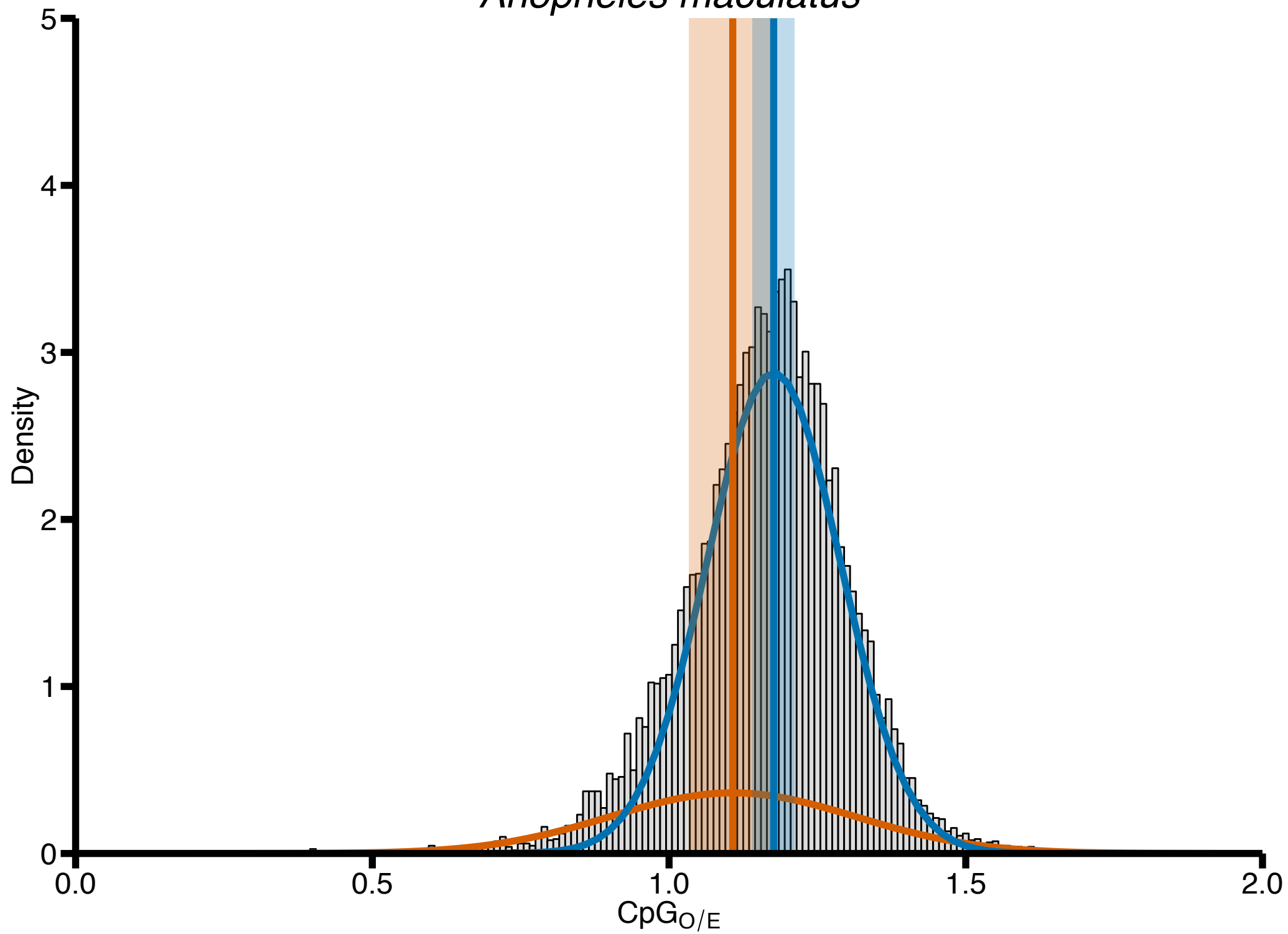
Anopheles melas
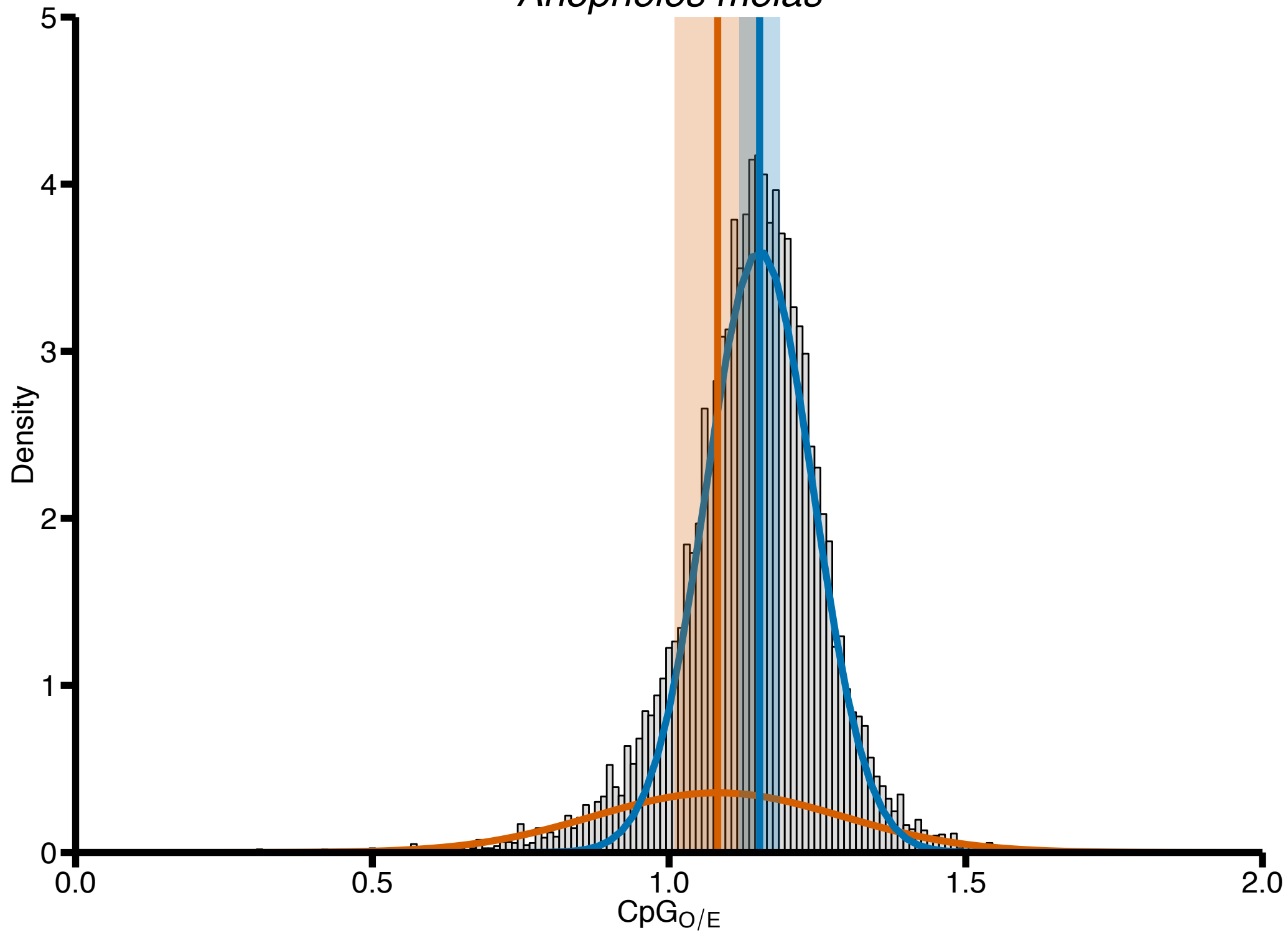
Anopheles merus
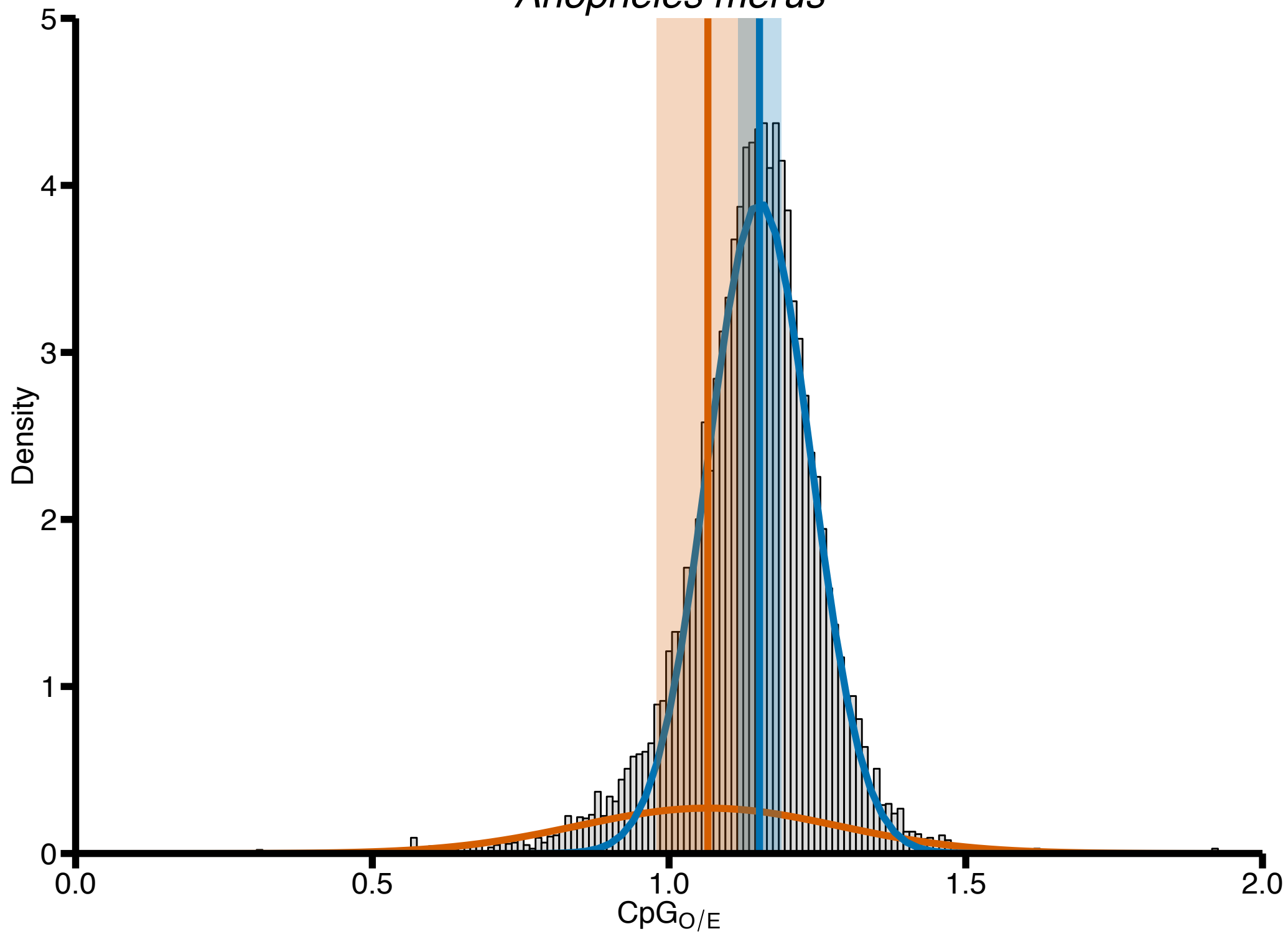
Anopheles minimus
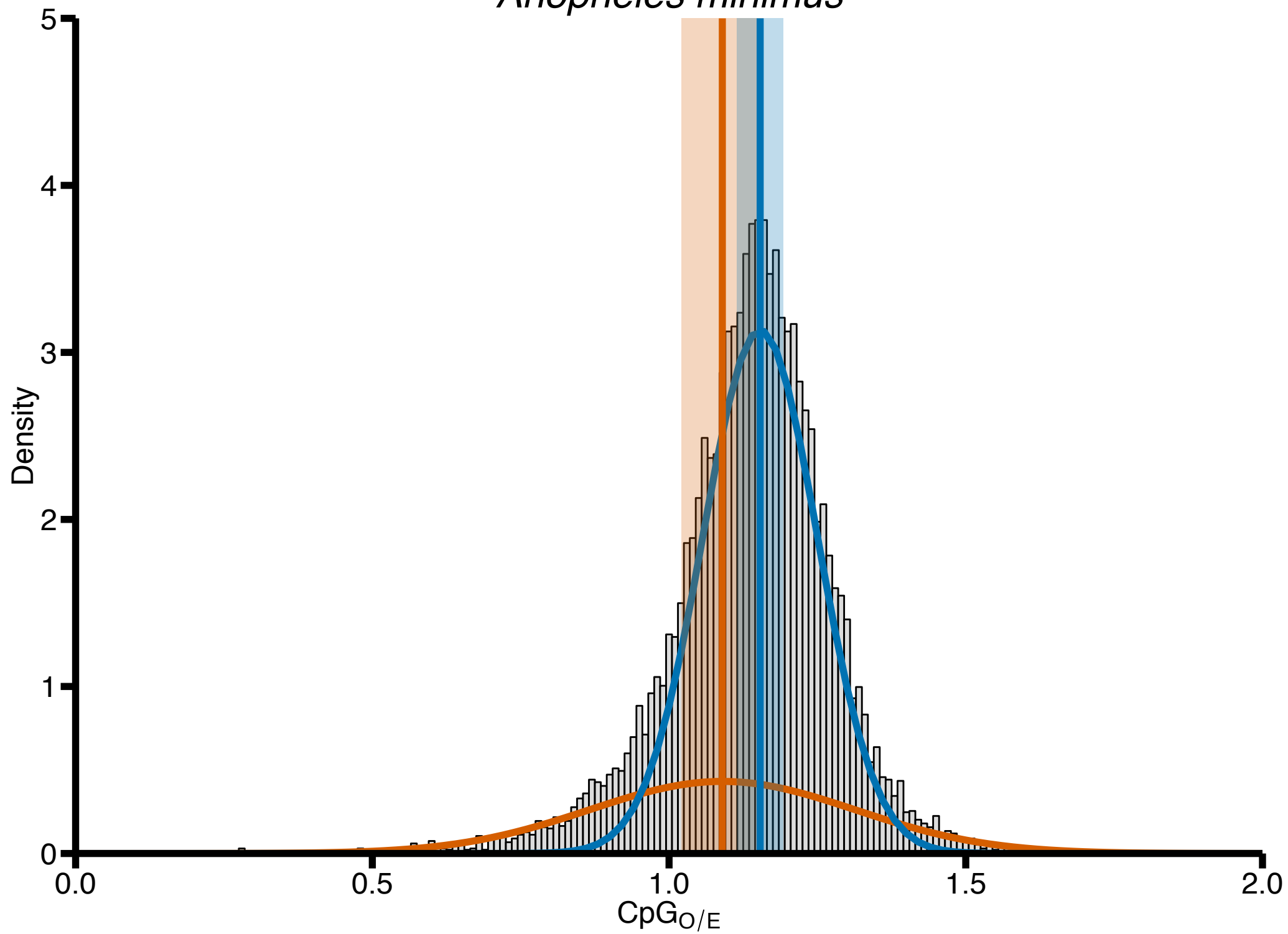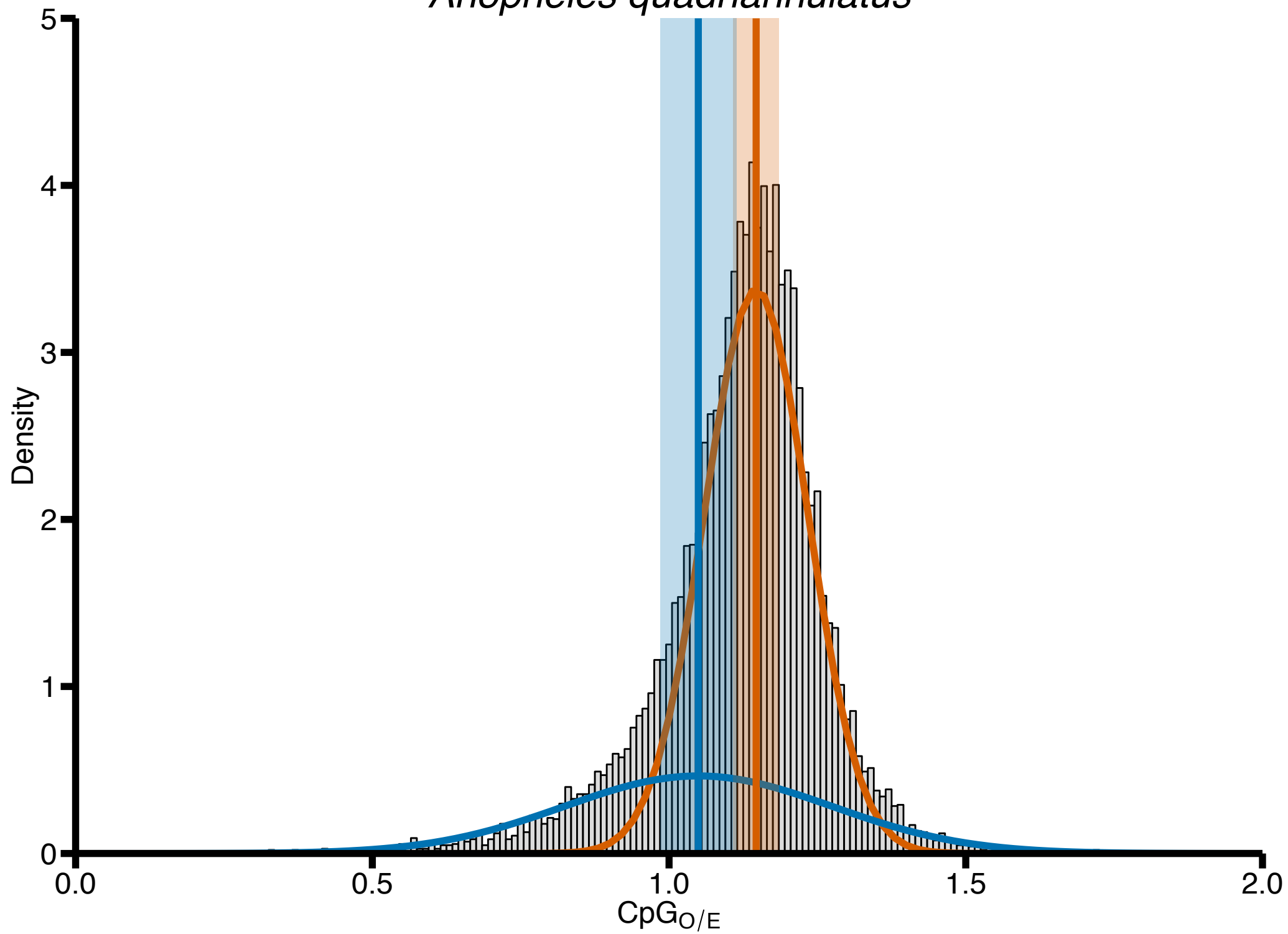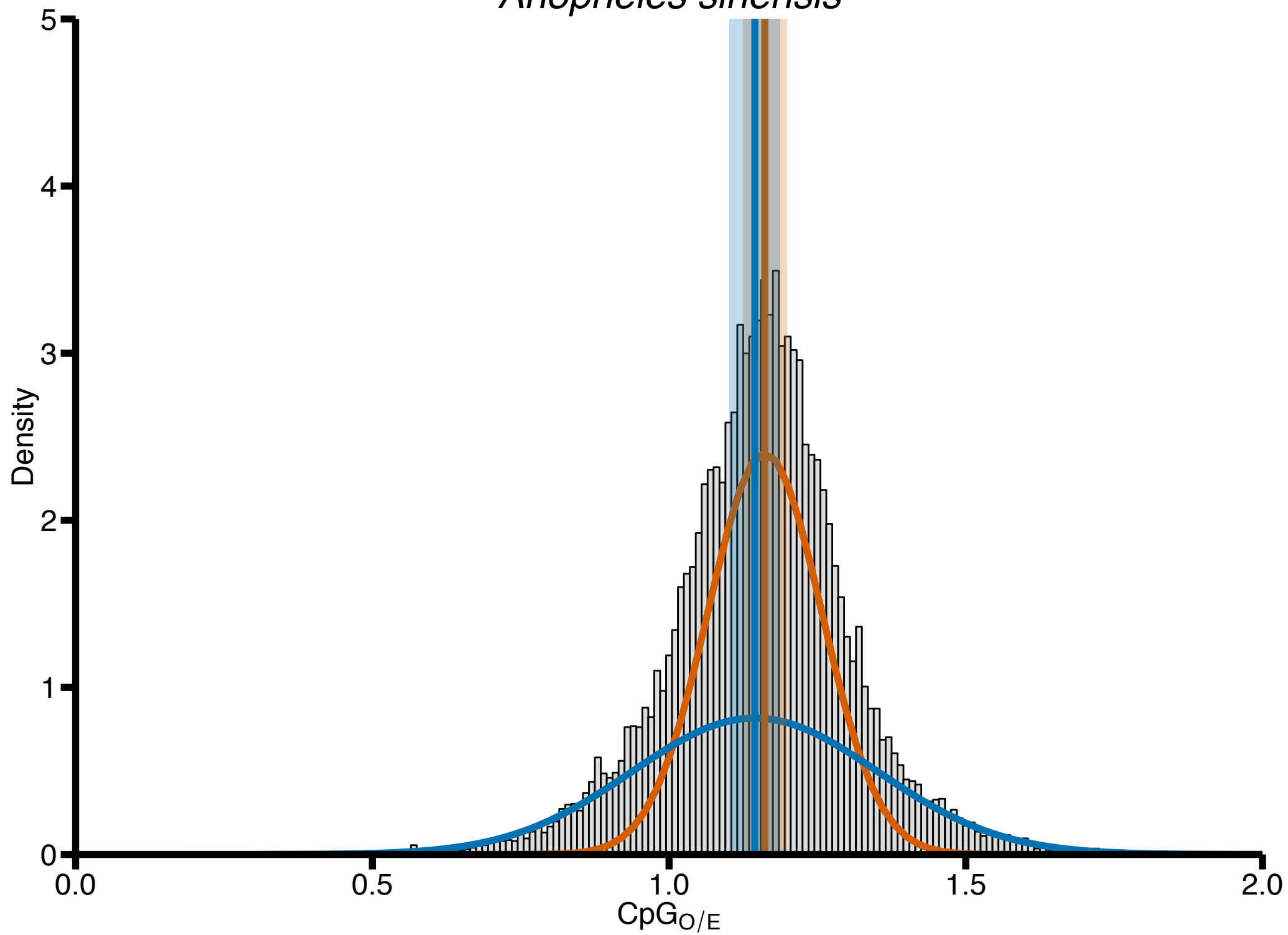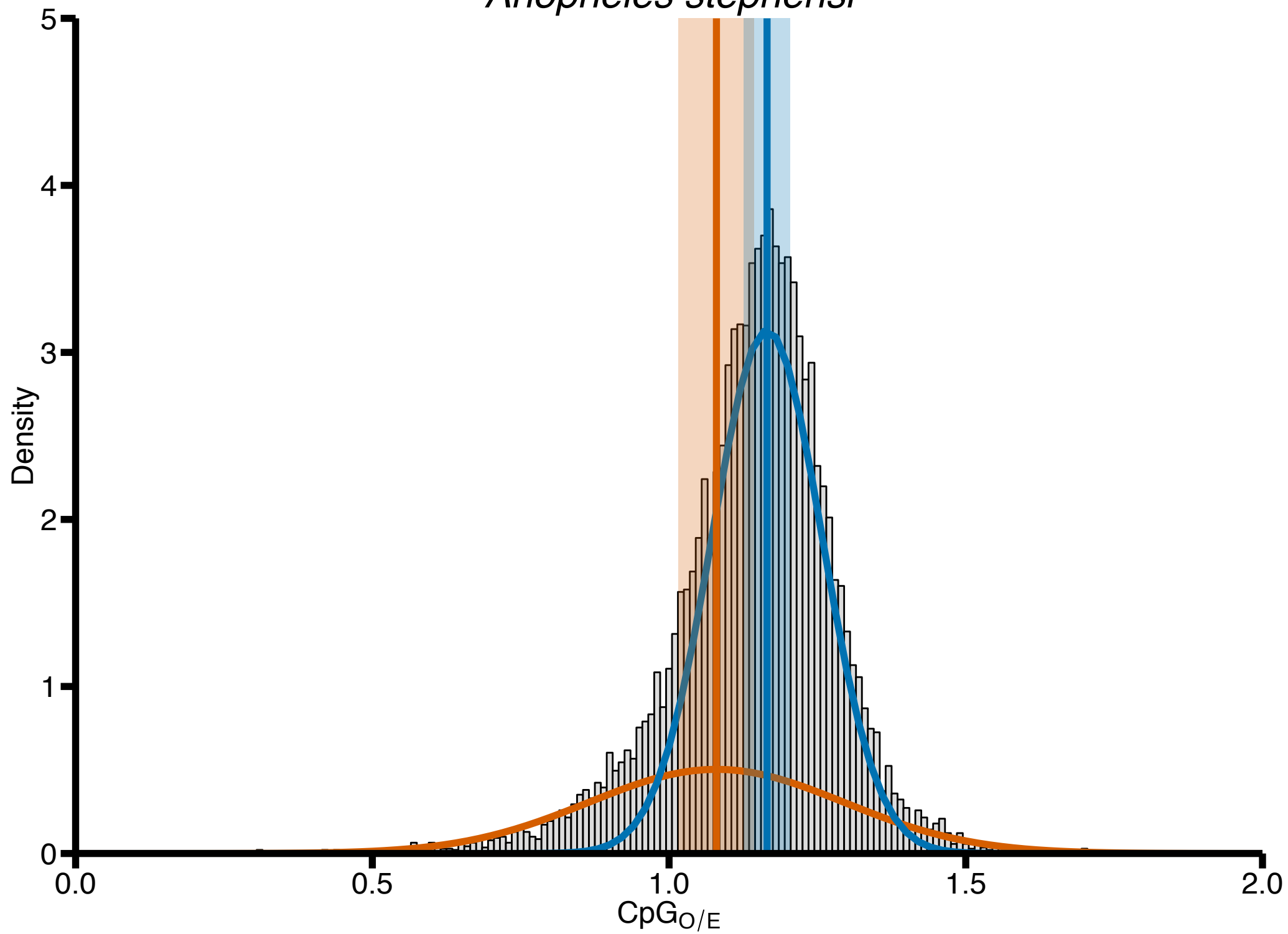
Bactrocera cucurbitae
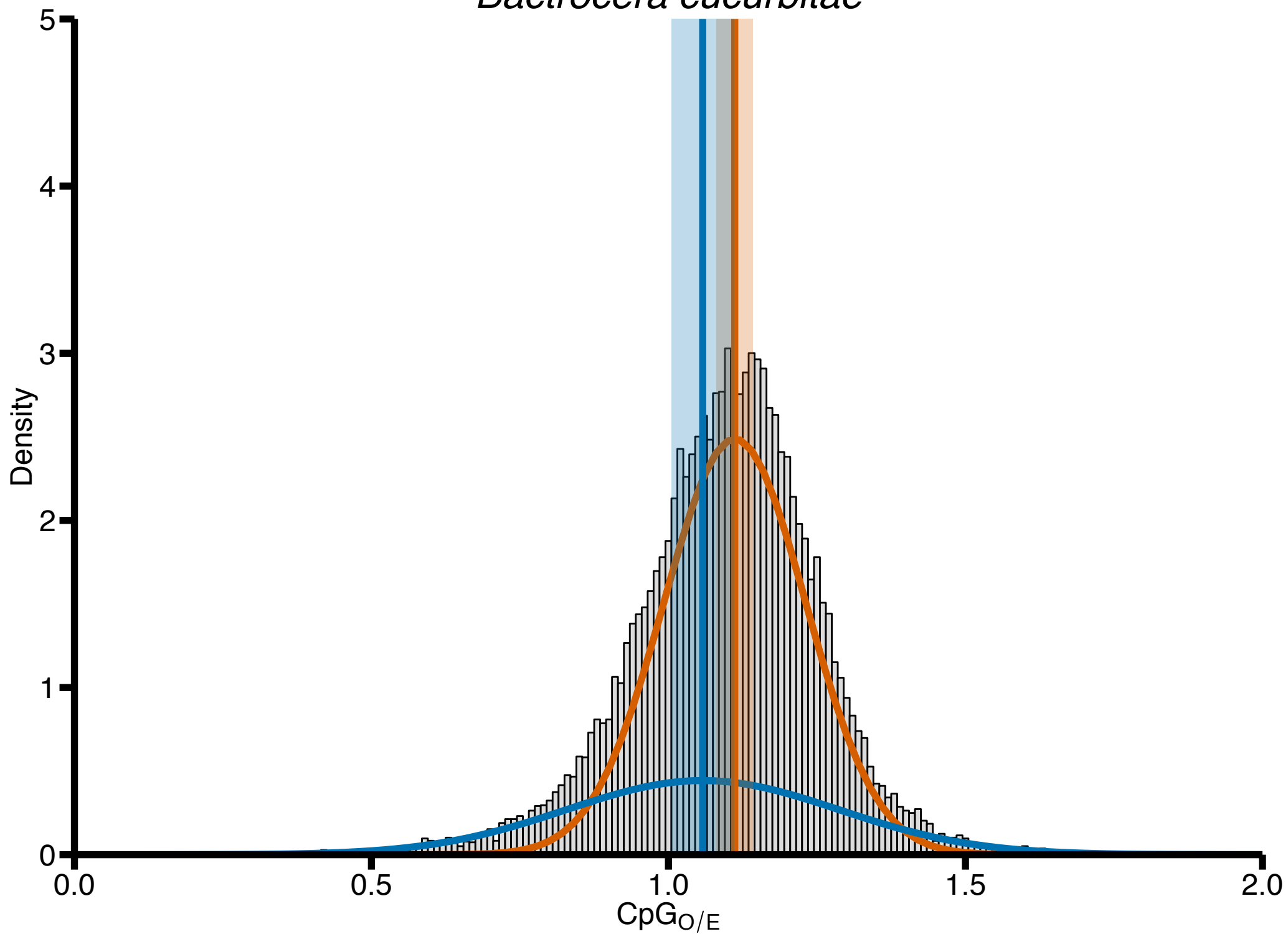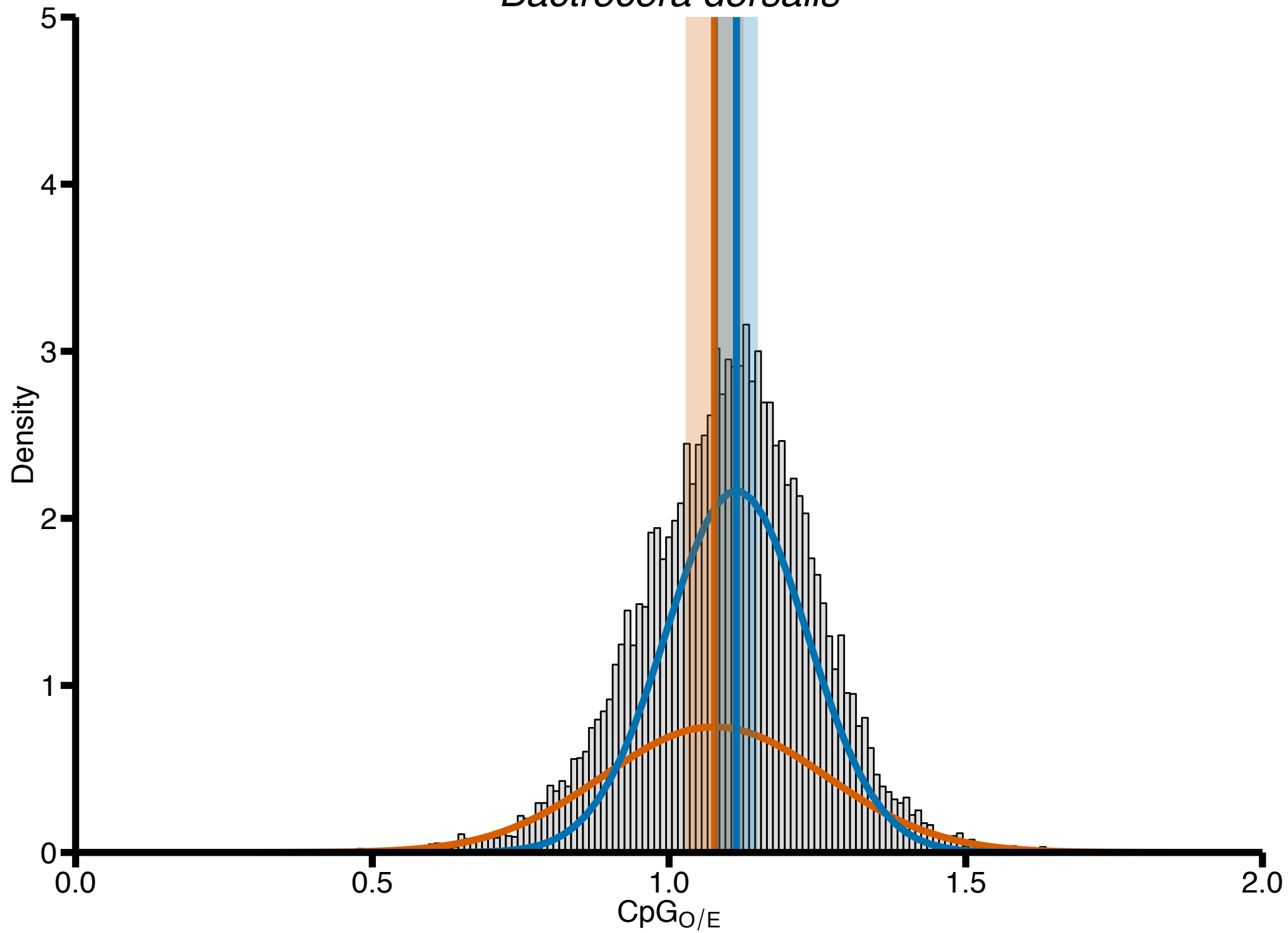
Bactrocera oleae
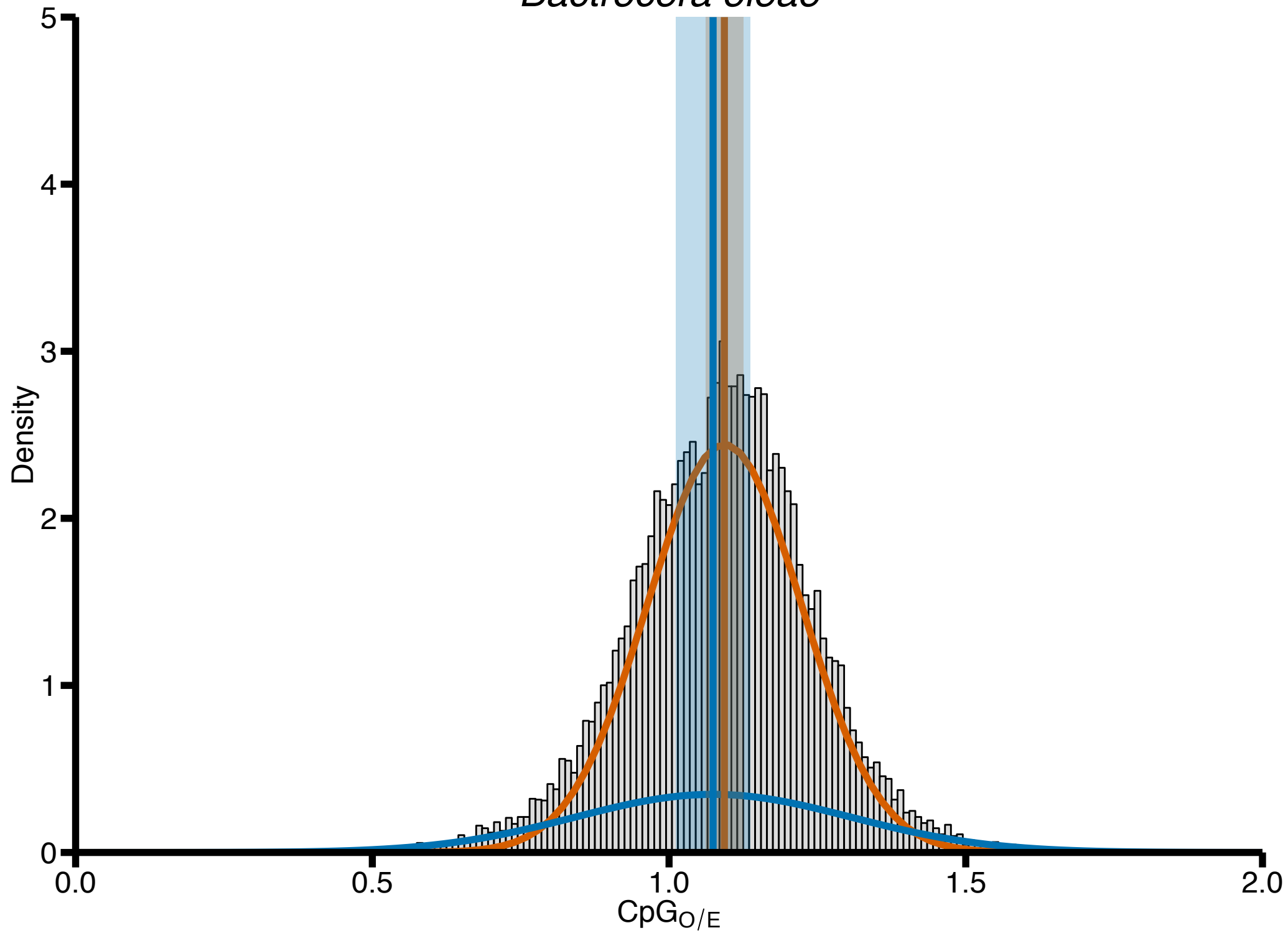
Ceratitis capitata

Density

CpG$_{O/E}$
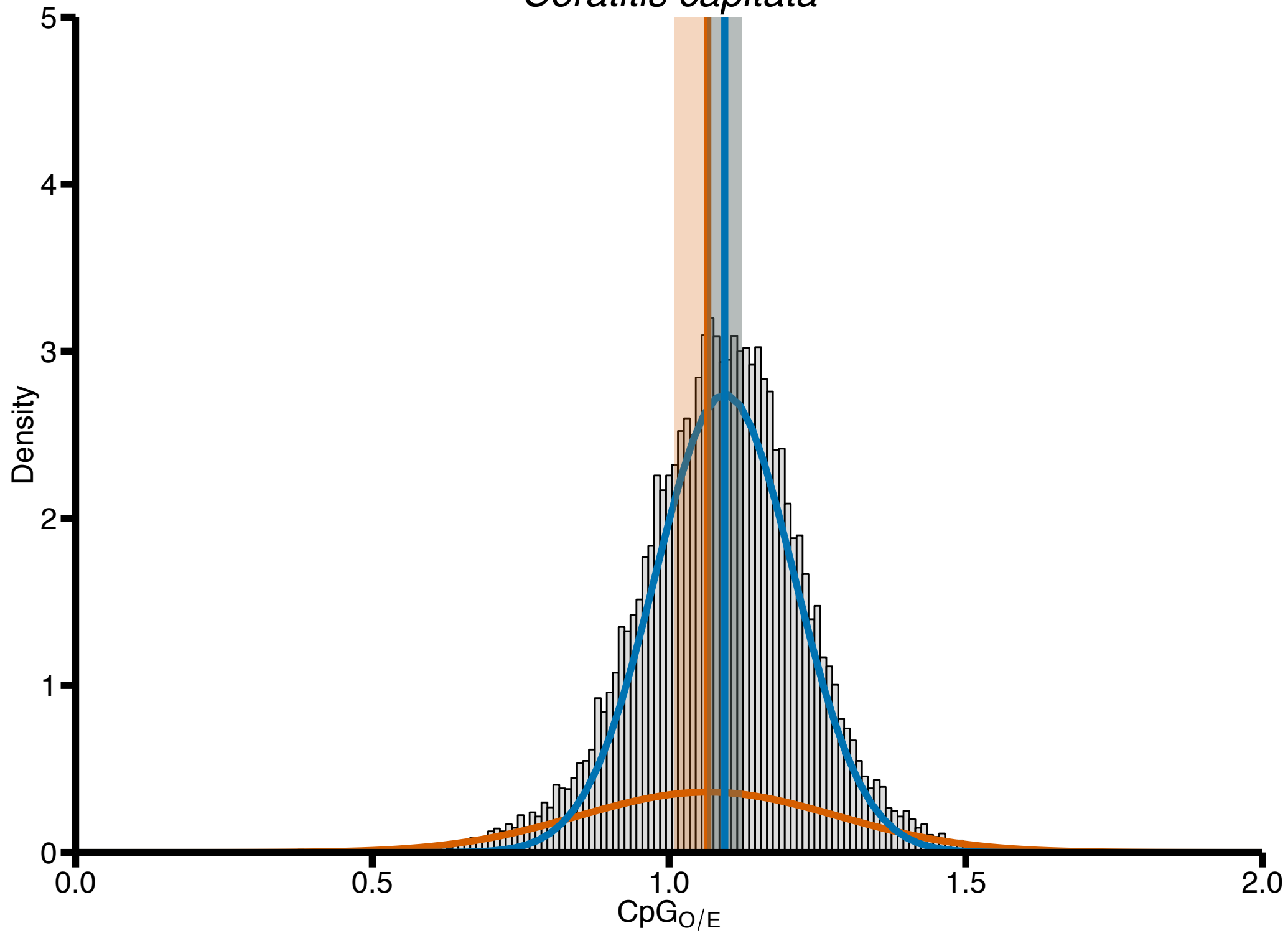
Culex quinquefasciatus
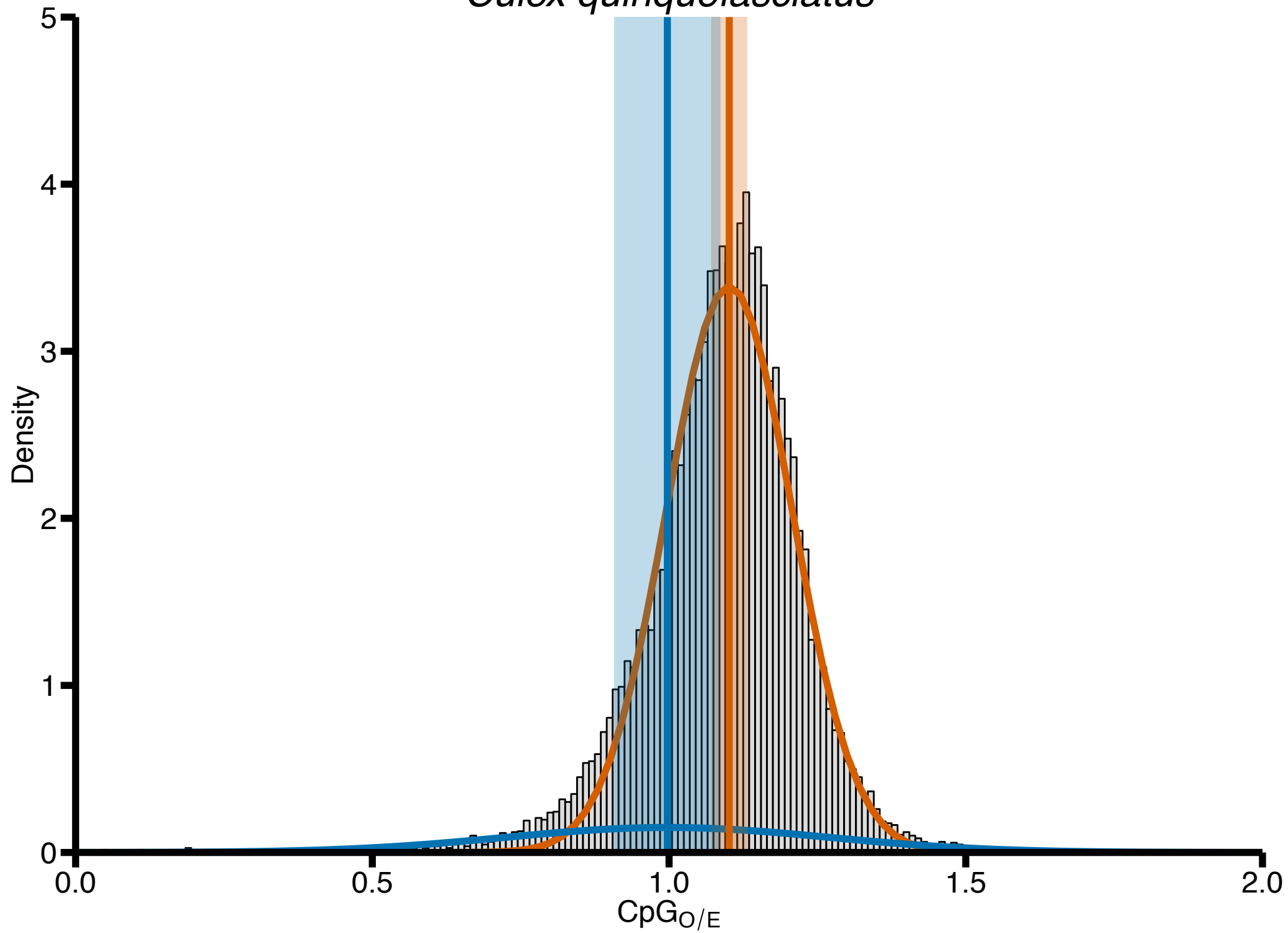
Drosophila ananassae
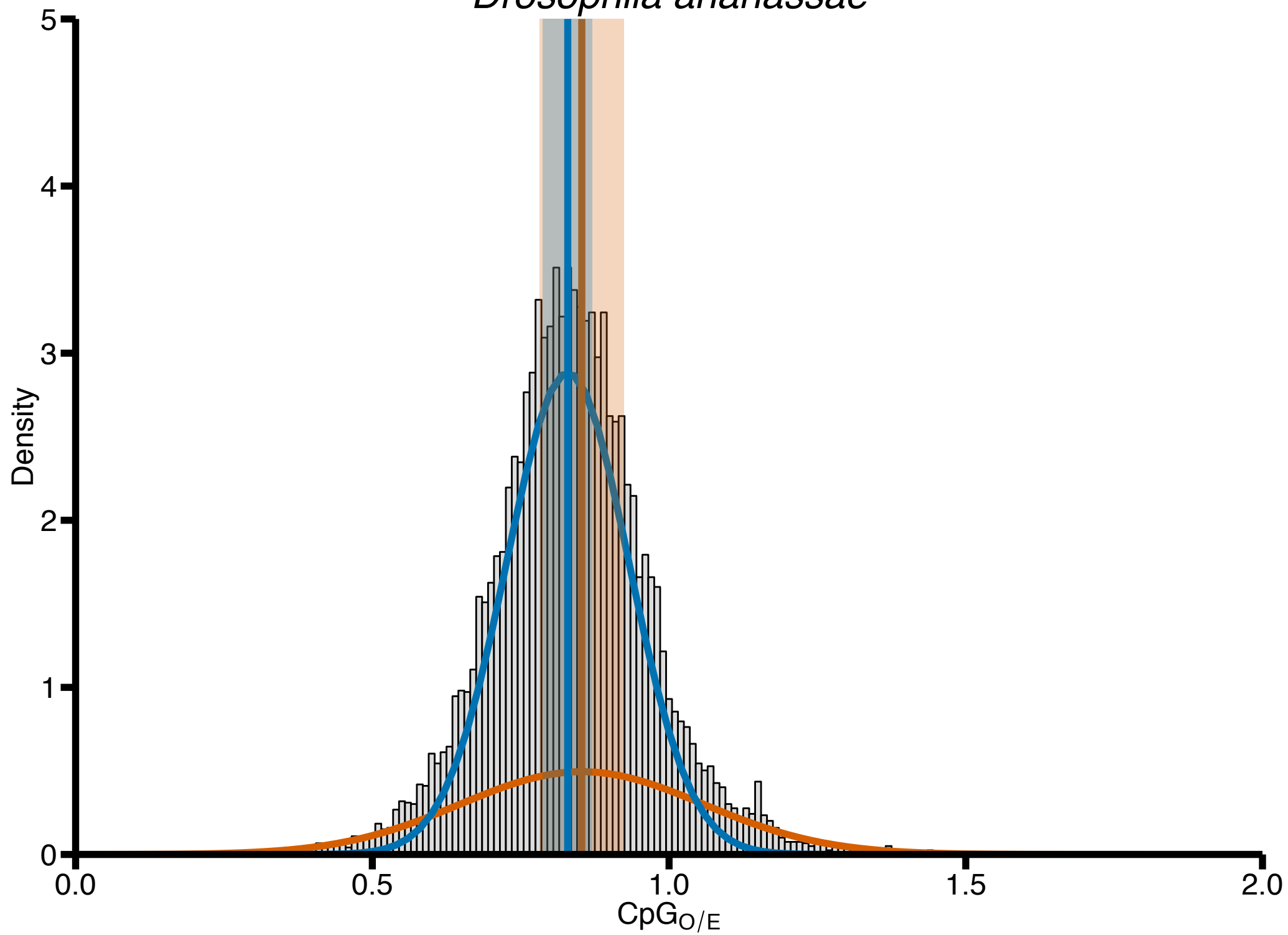
Drosophila biarmipes
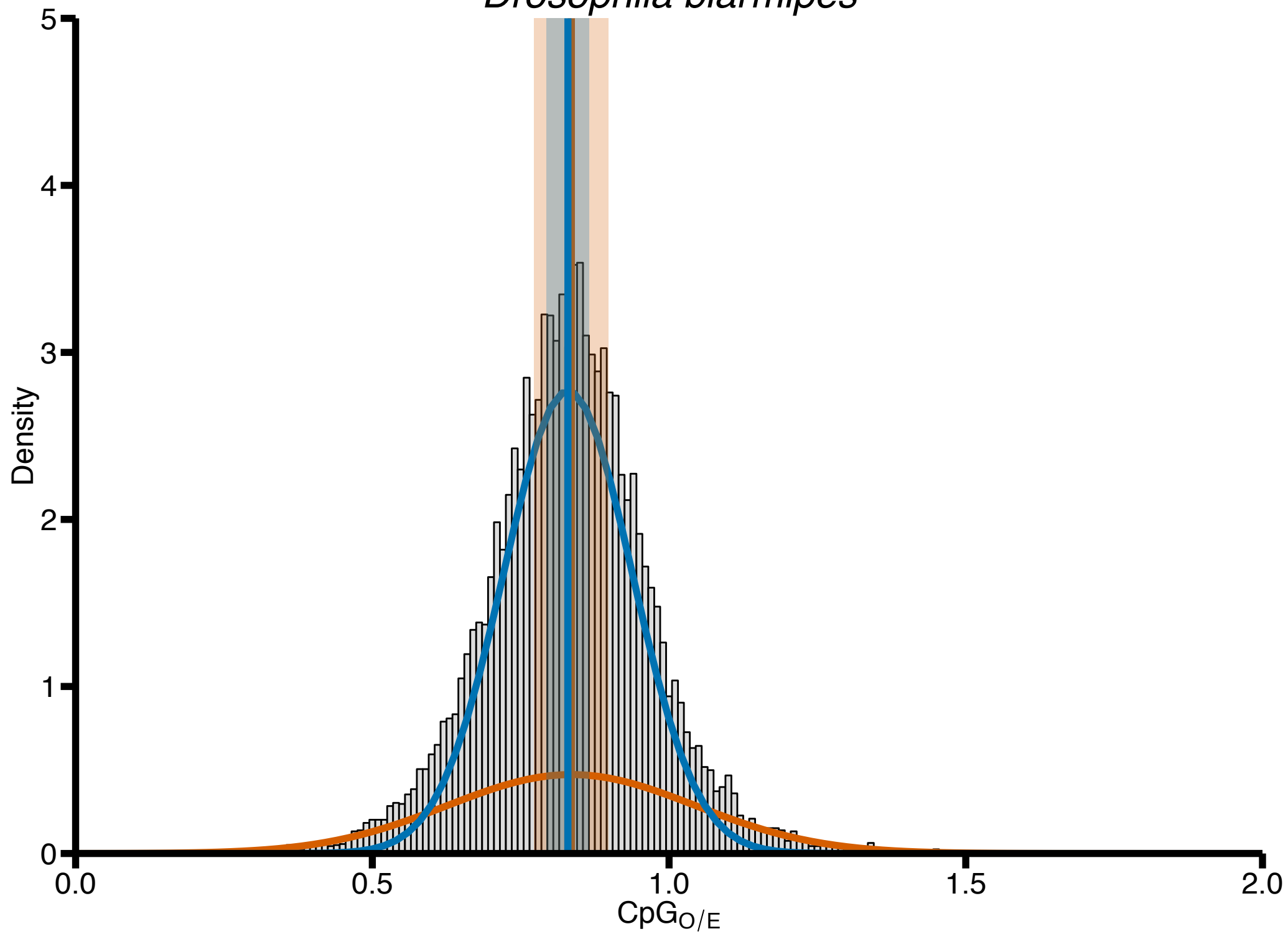
Drosophila bipectinata
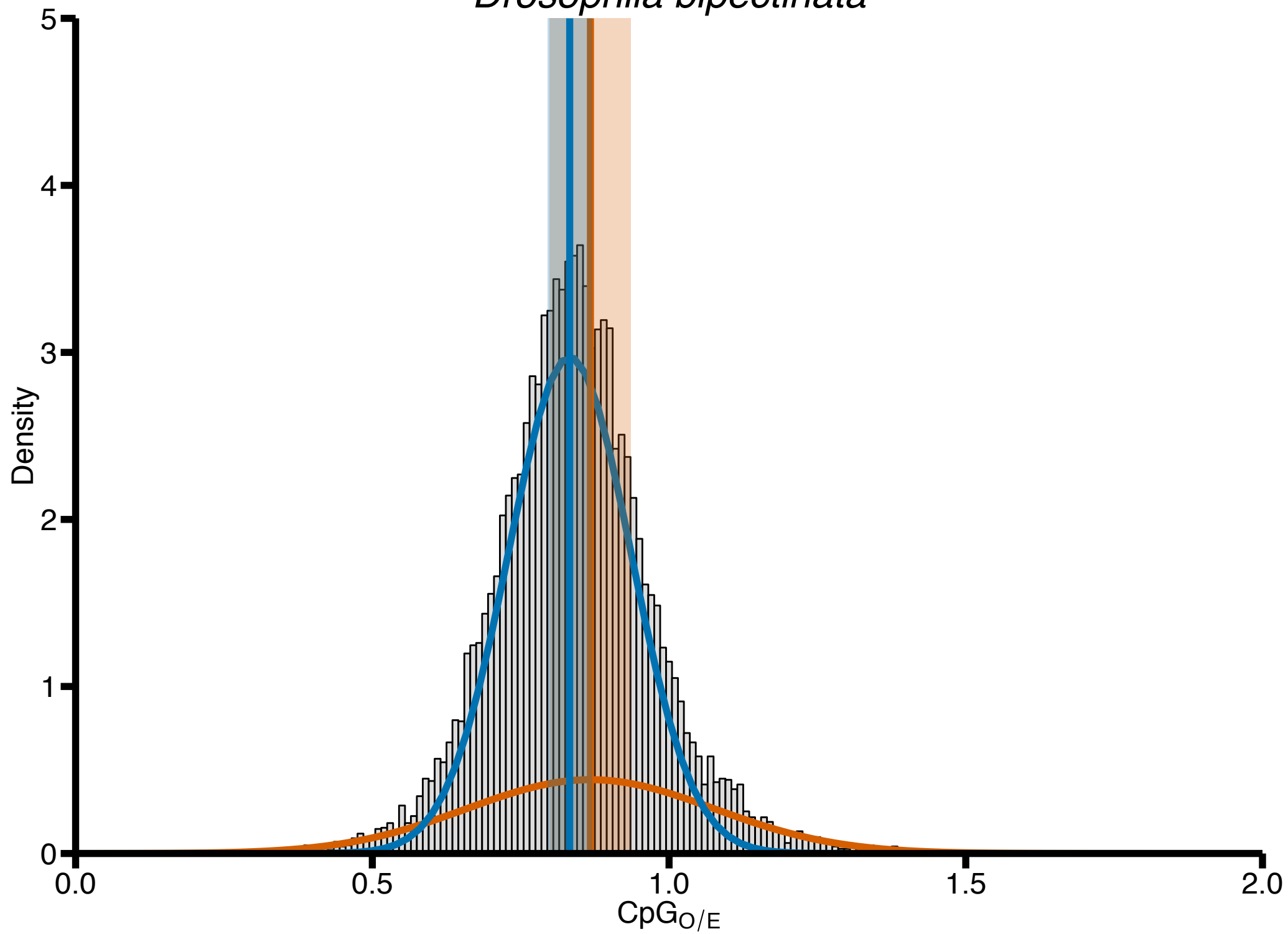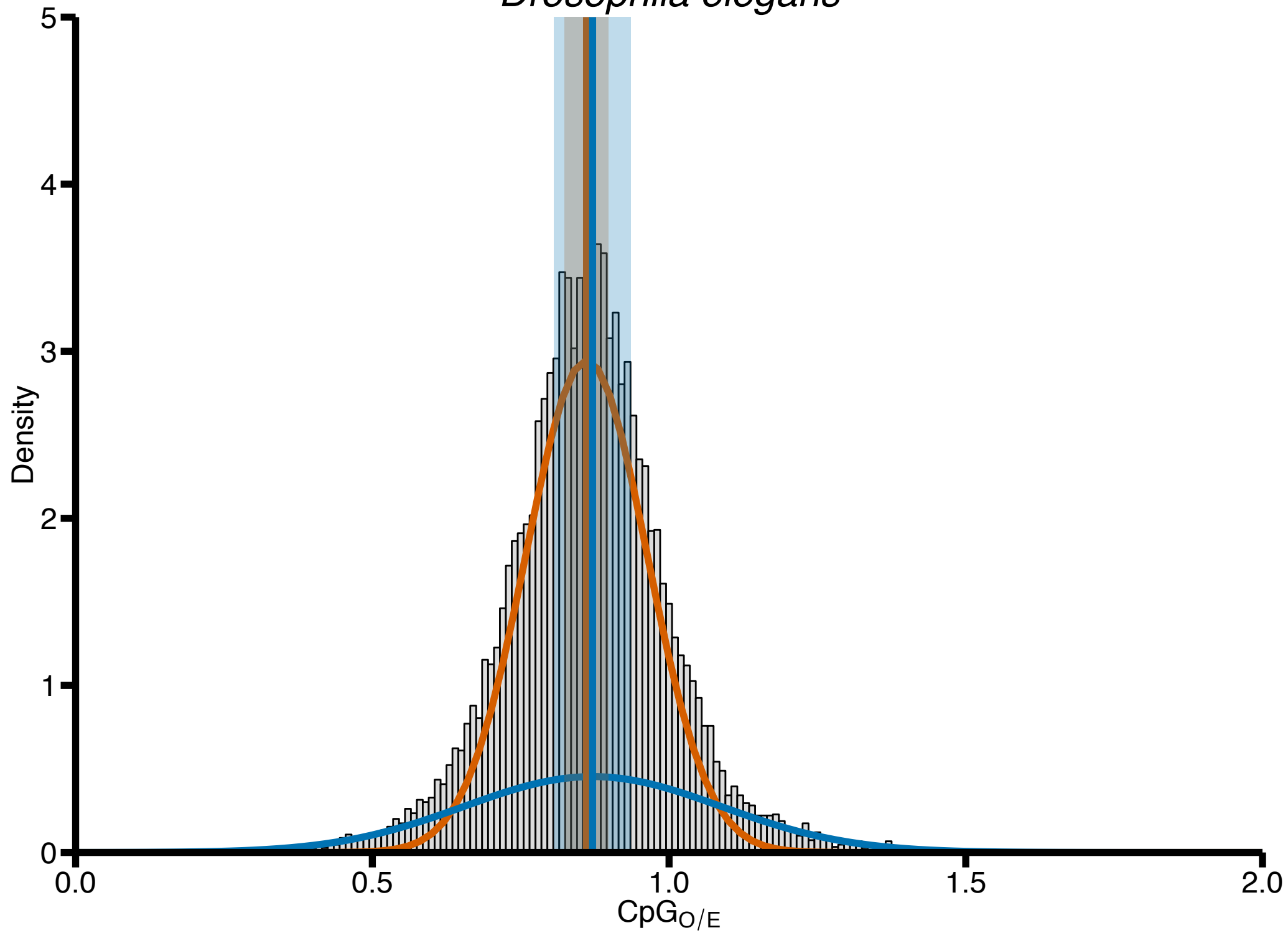
Drosophila erecta
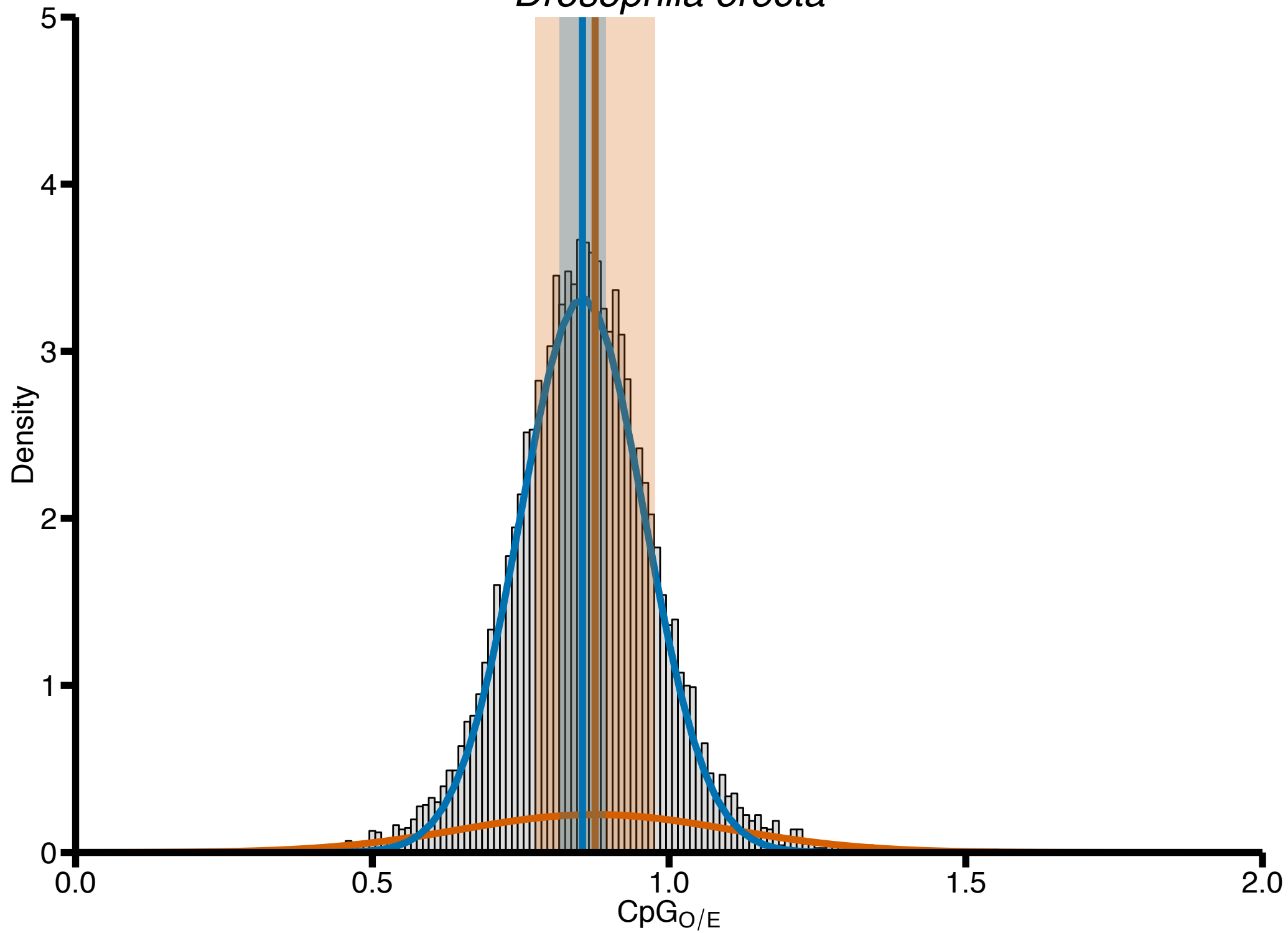
Drosophila eugracilis
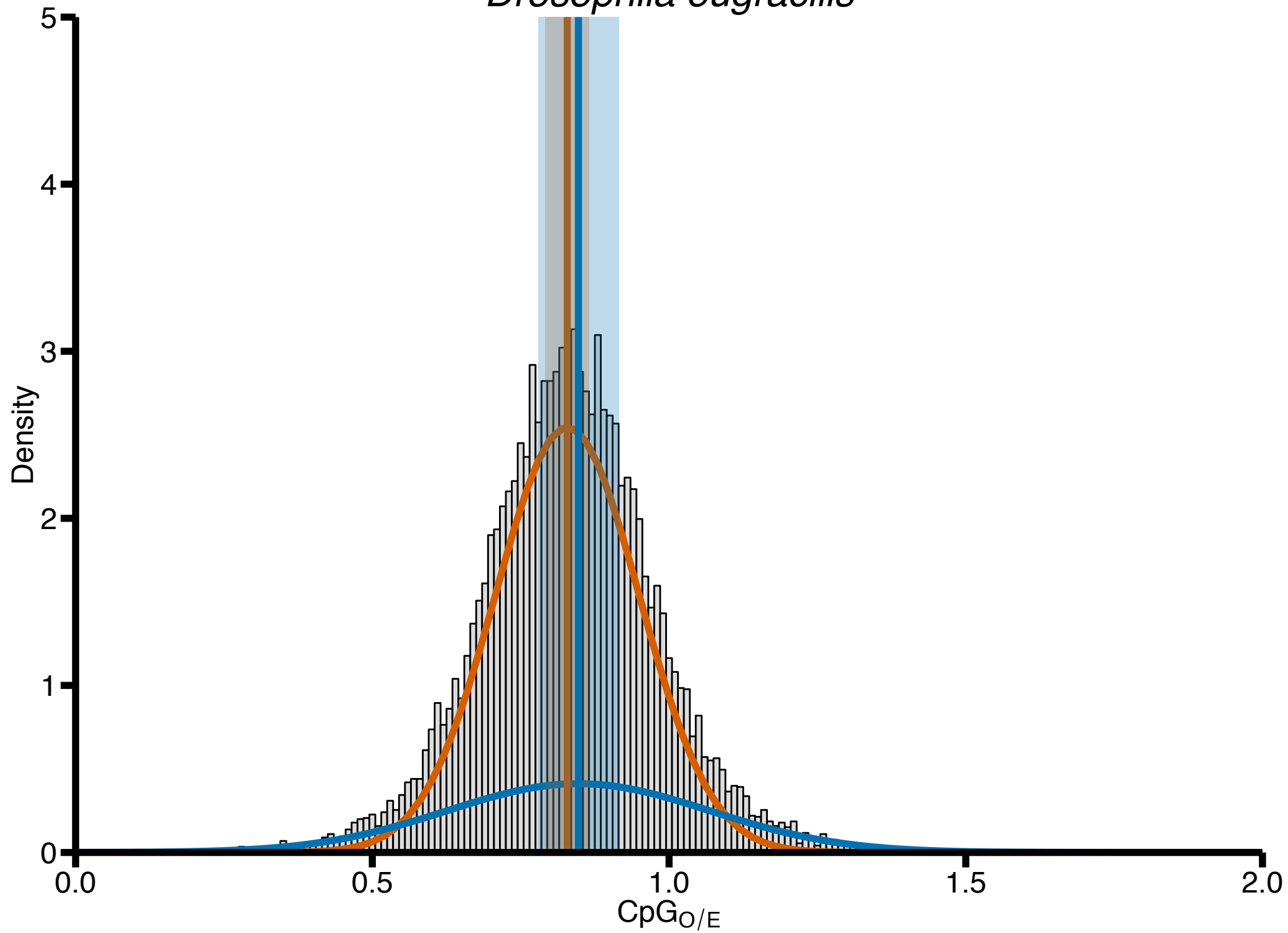
Drosophila ficusphila
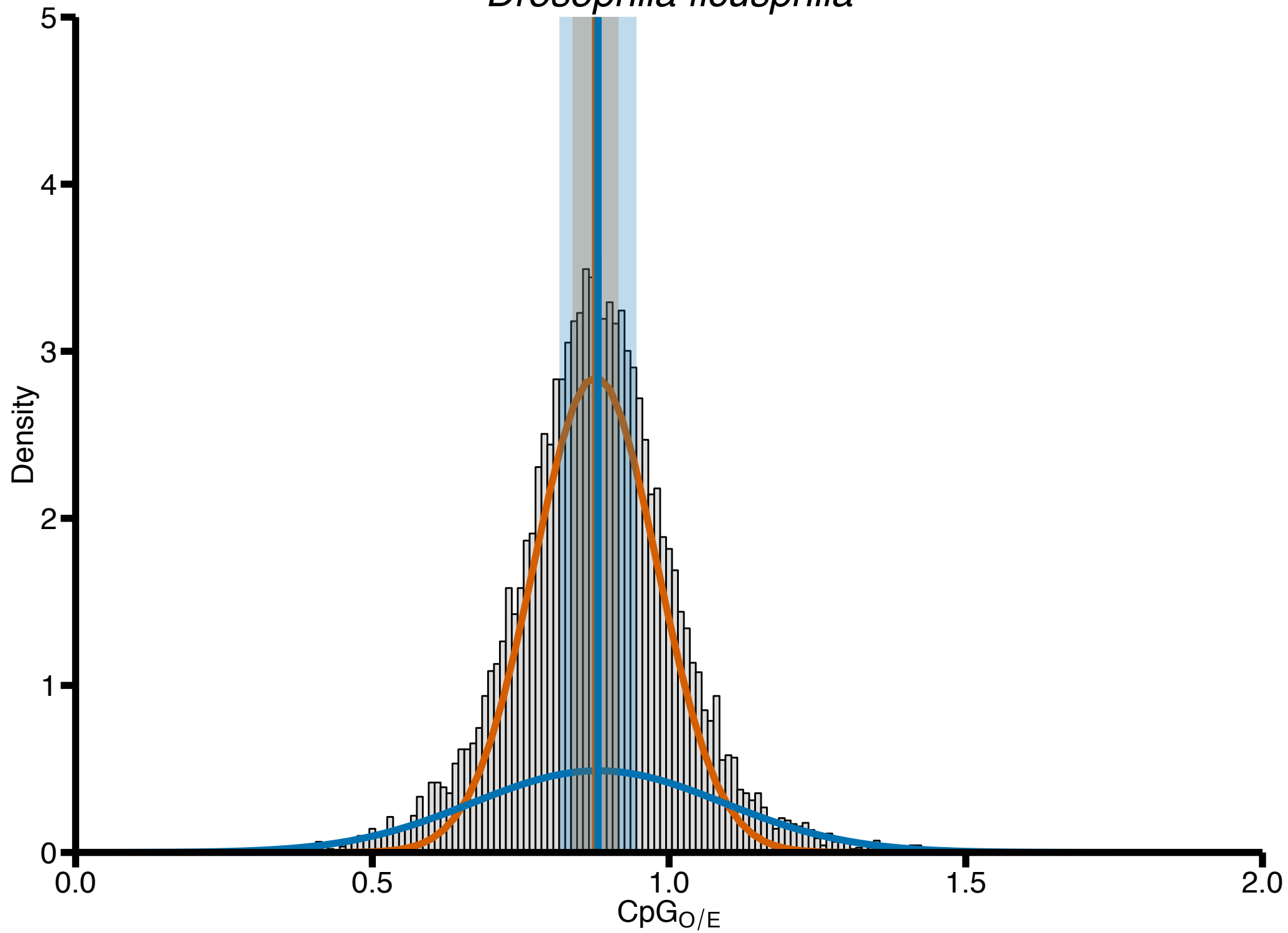
Drosophila grimshawi
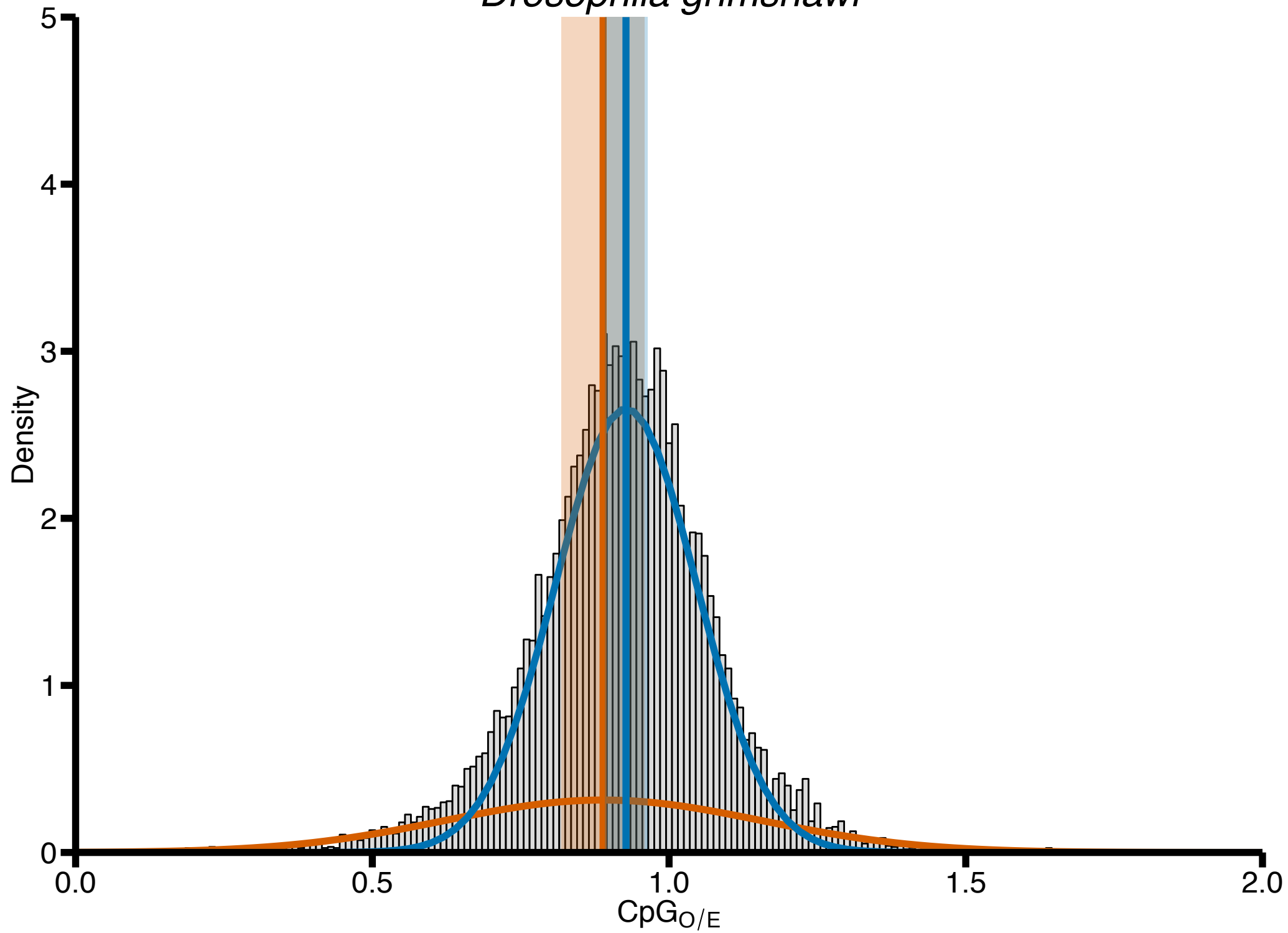
Drosophila kikkawai
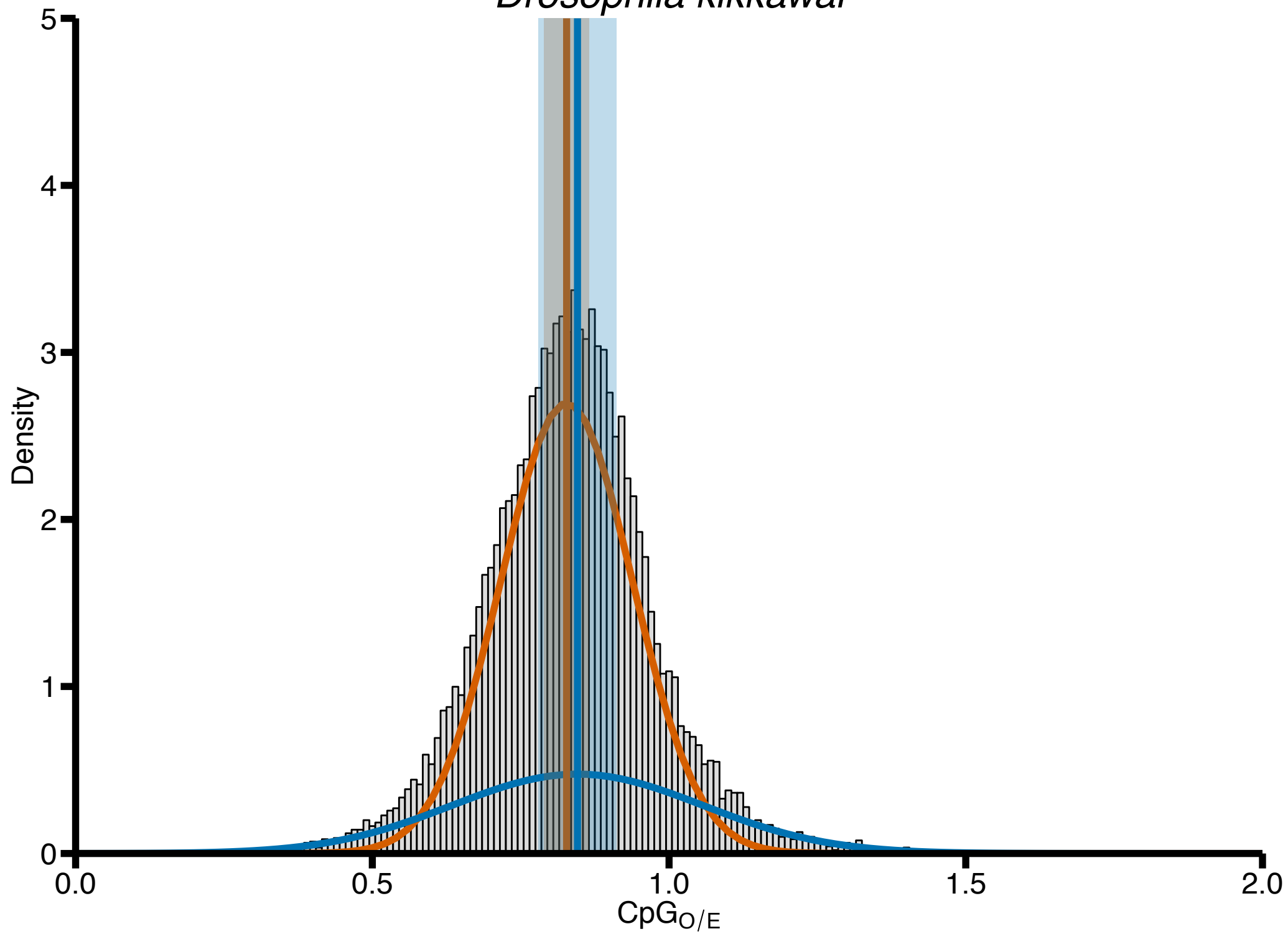
Drosophila melanogaster
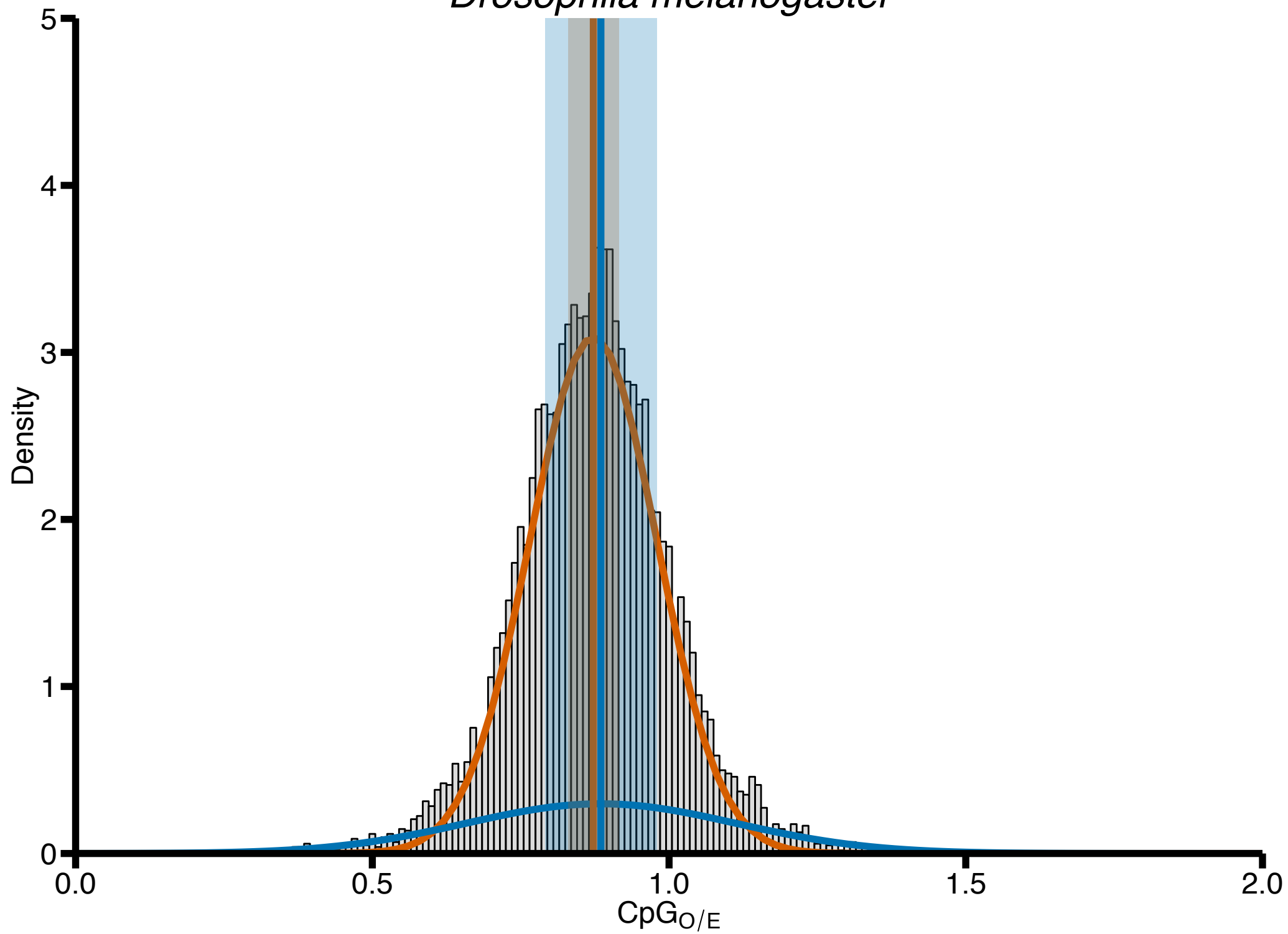
Drosophila mojavensis
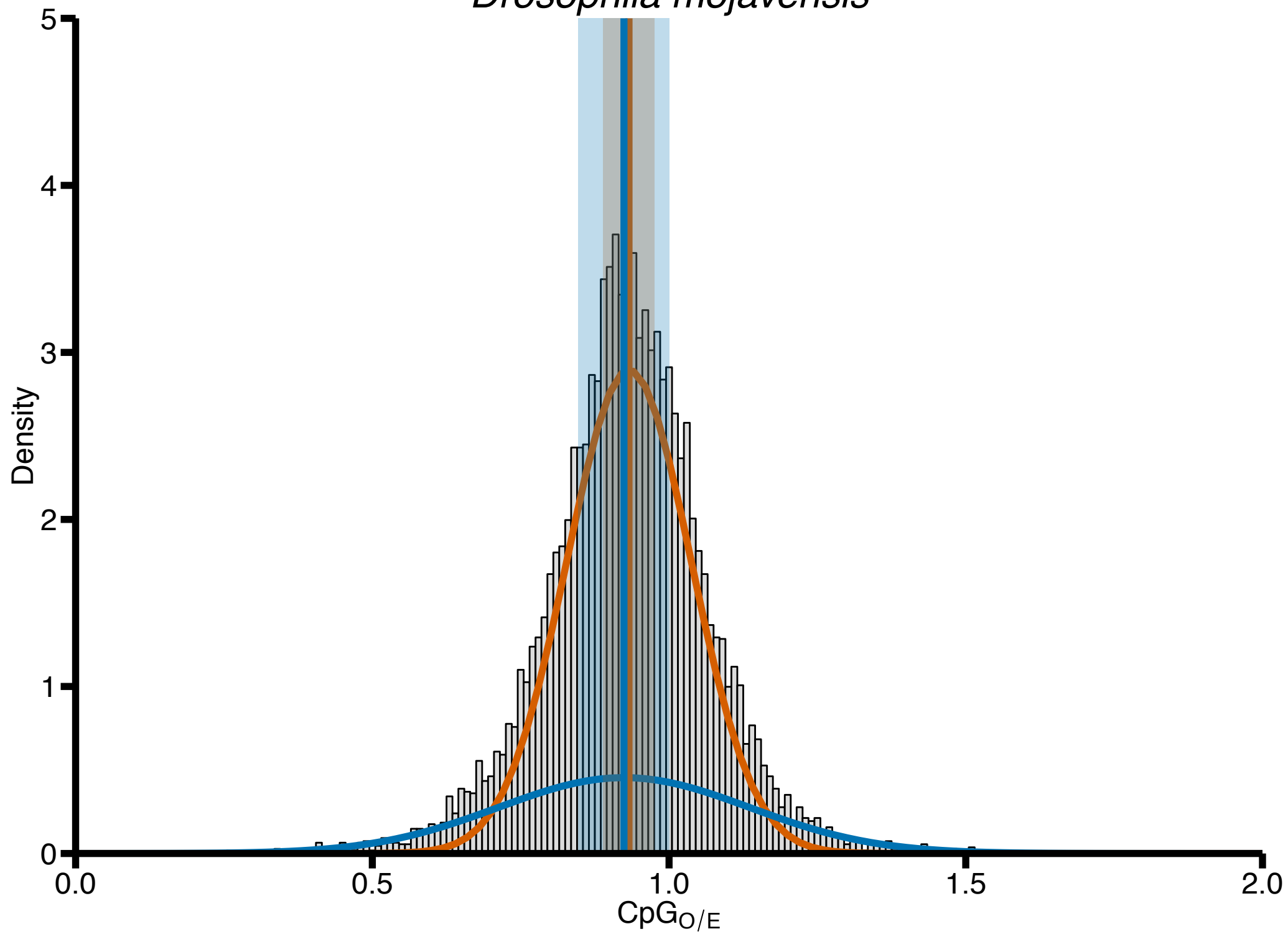
Drosophila persimilis
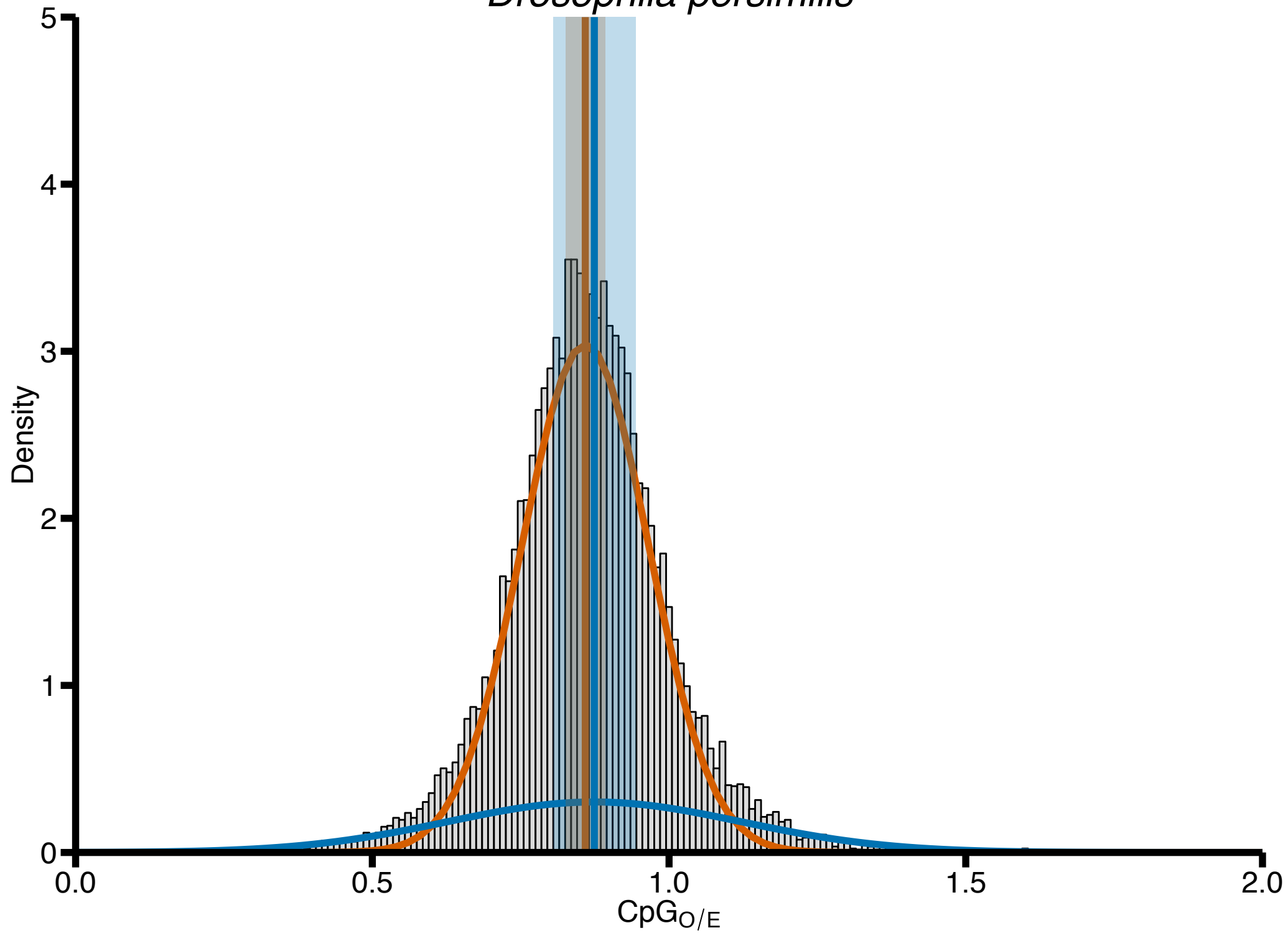
Drosophila pseudoobscura
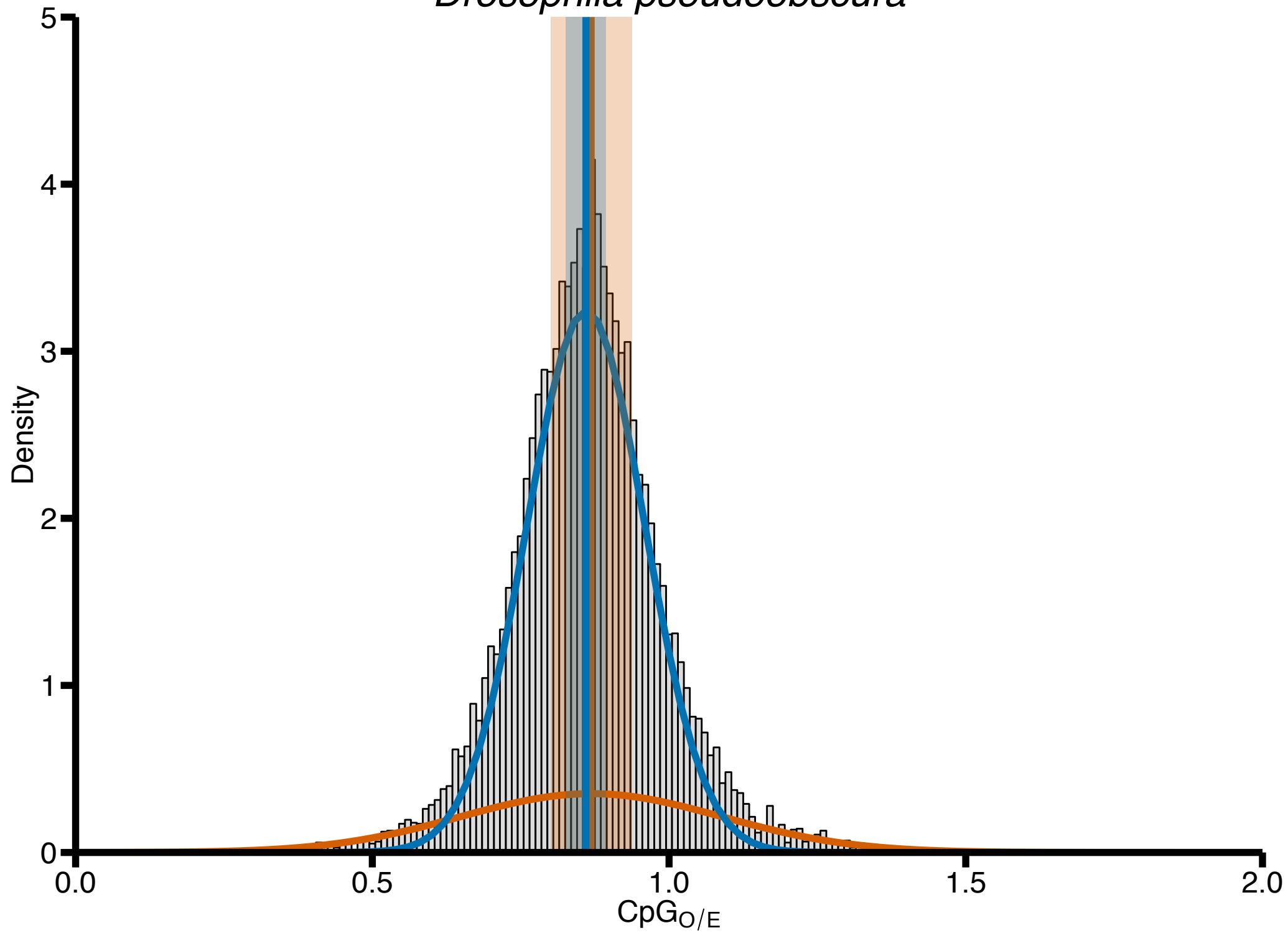
Drosophila rhopaloa
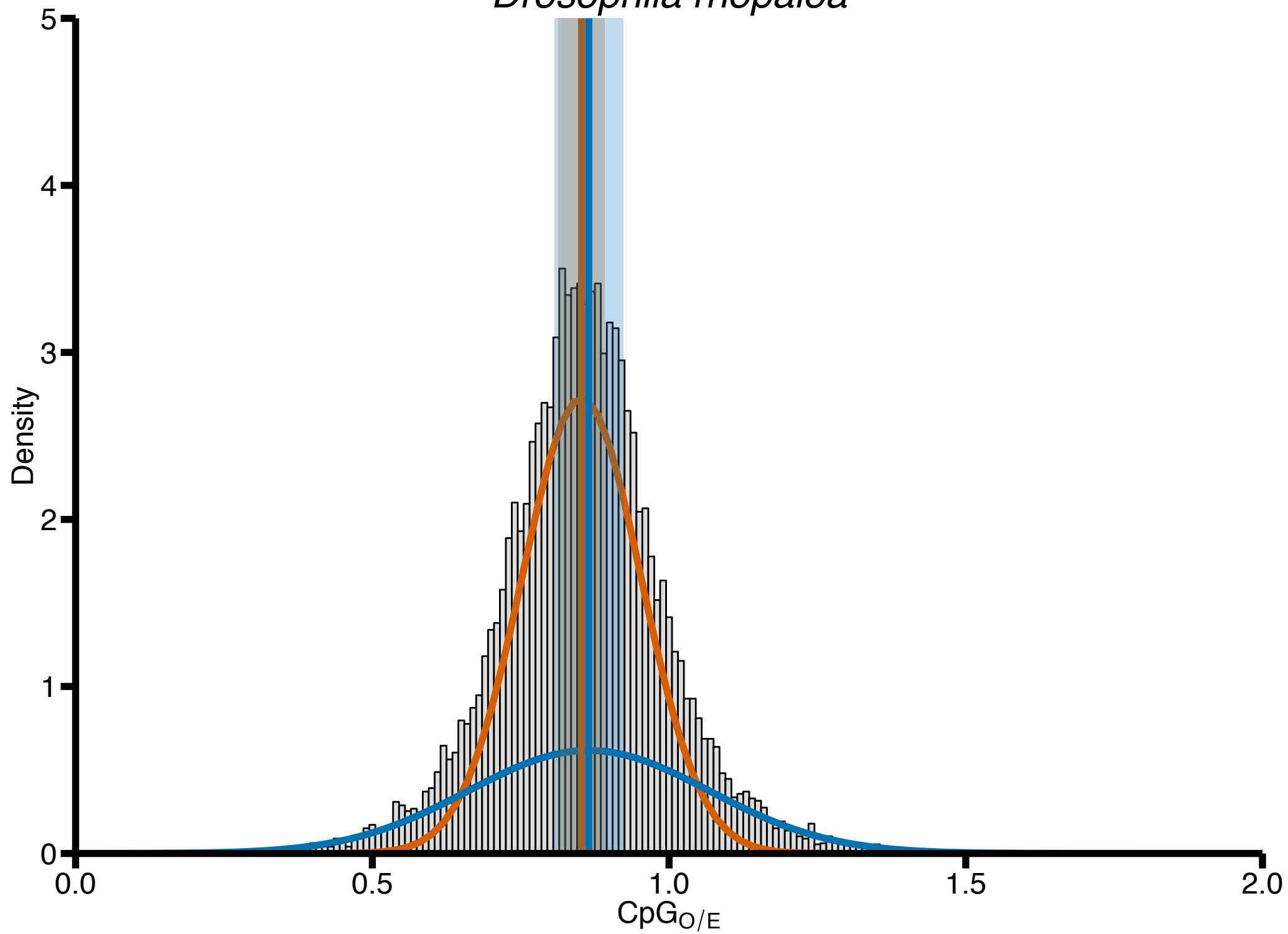
Drosophila sechellia
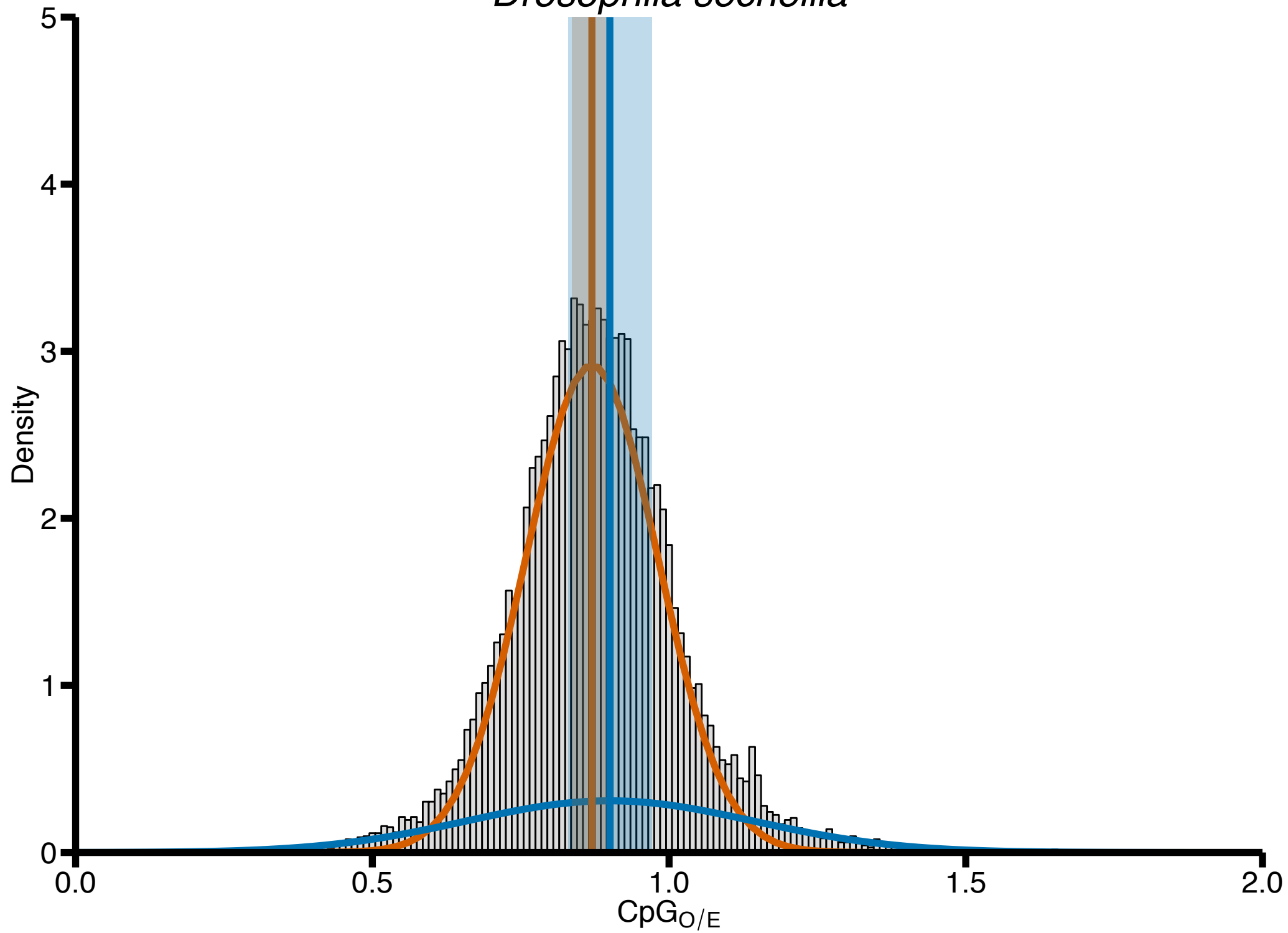
Drosophila simulans
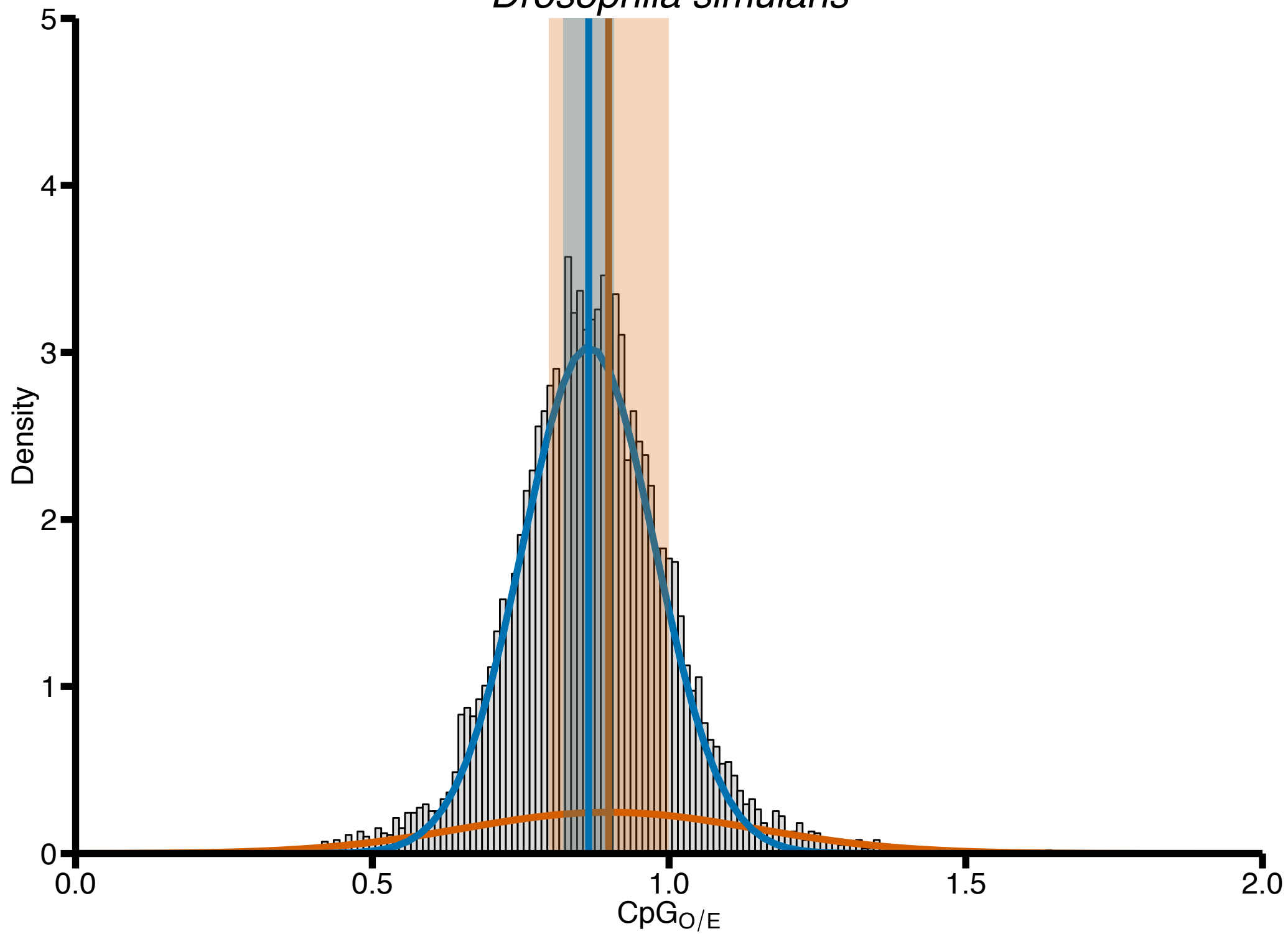
Drosophila suzukii
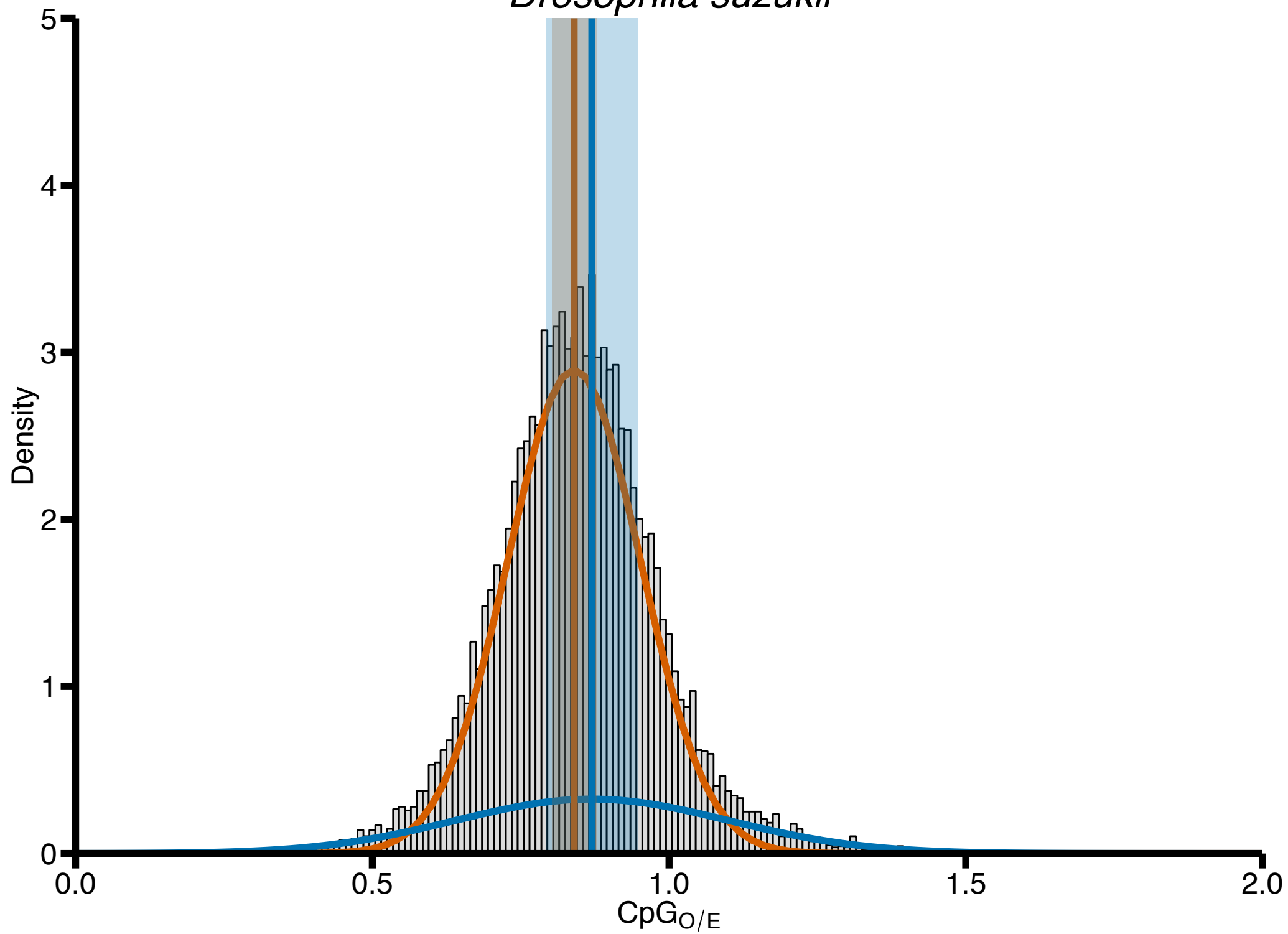
Drosophila takahashii
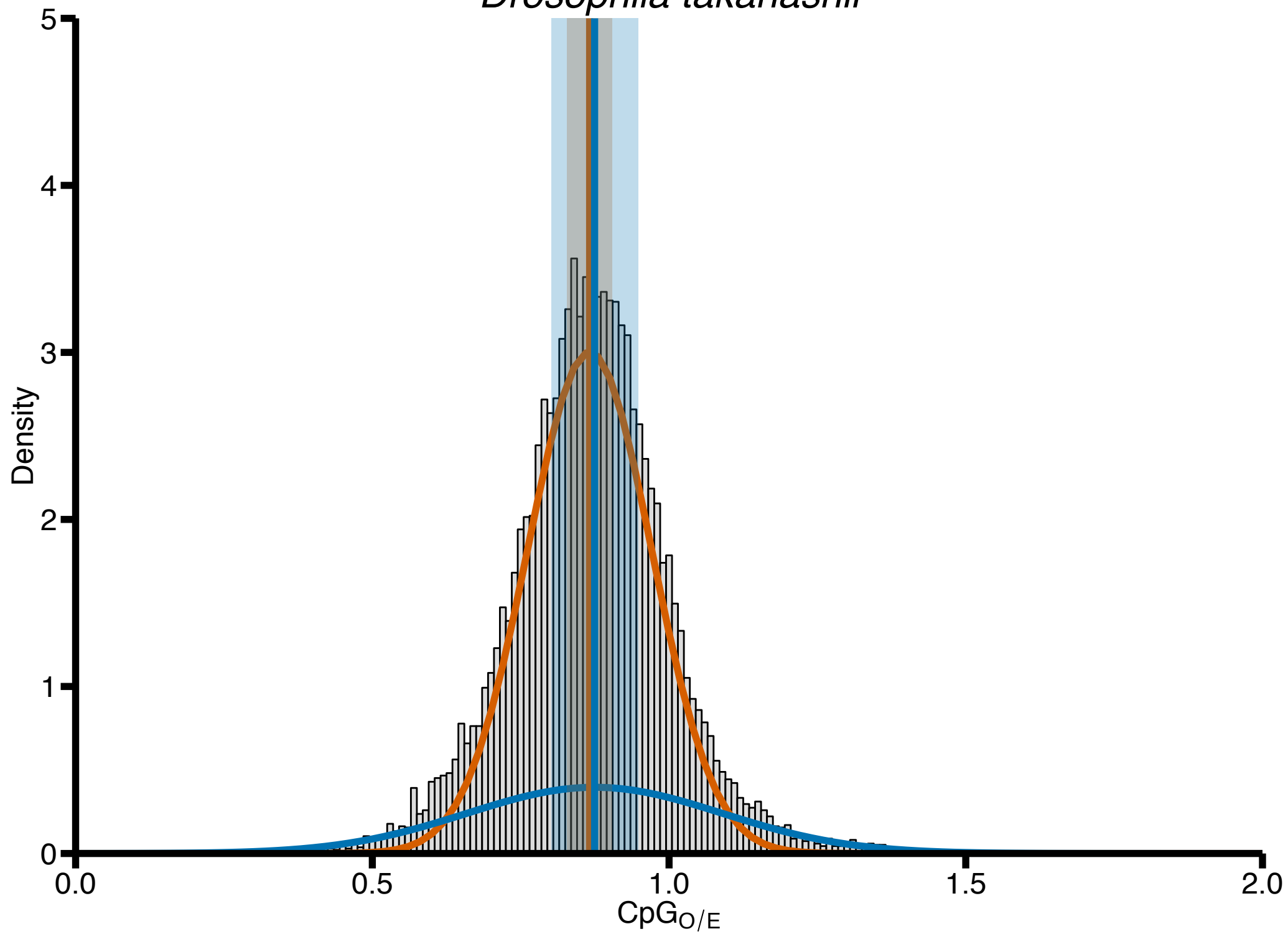
Drosophila virilis
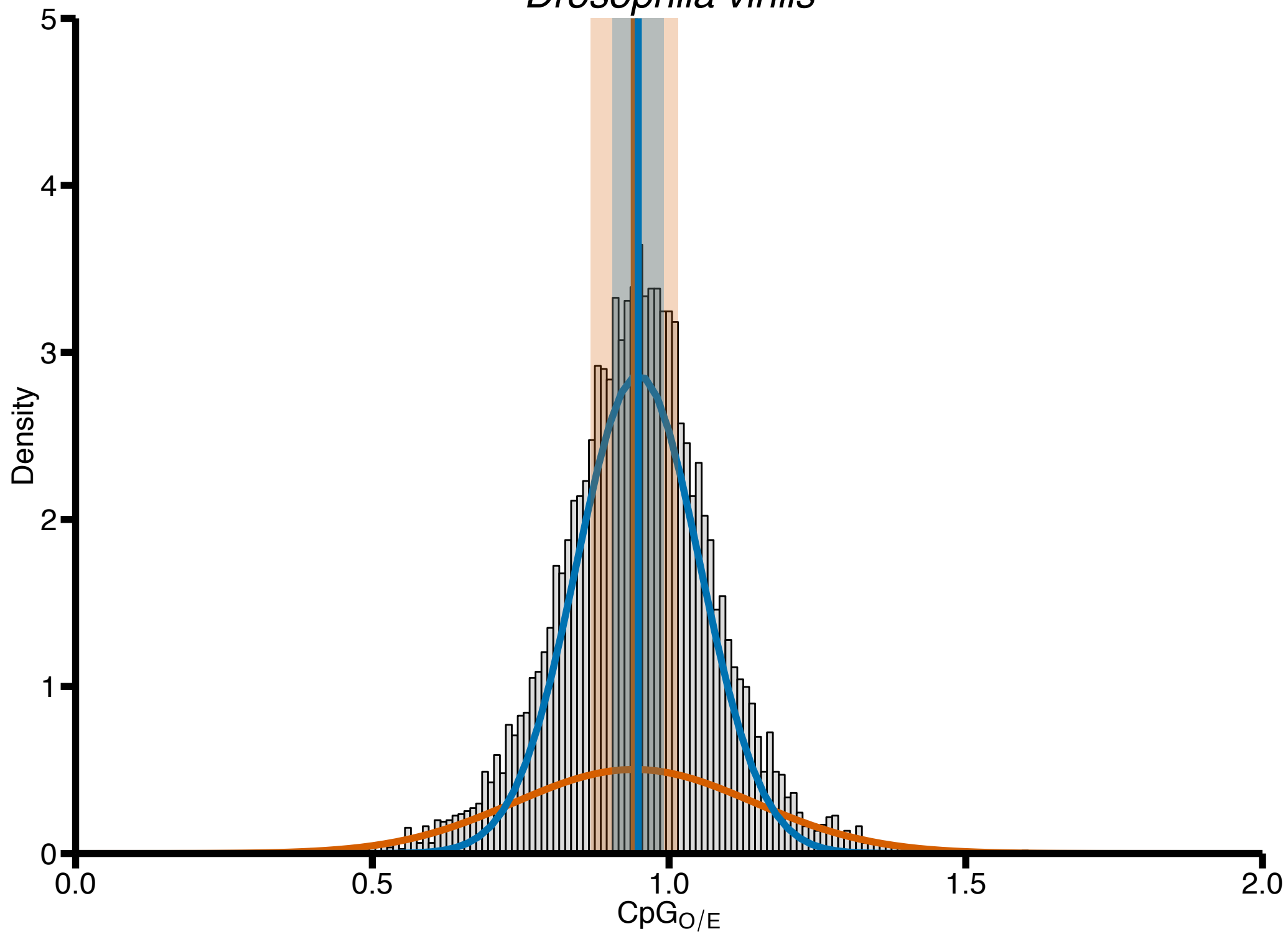
Drosophila willistoni
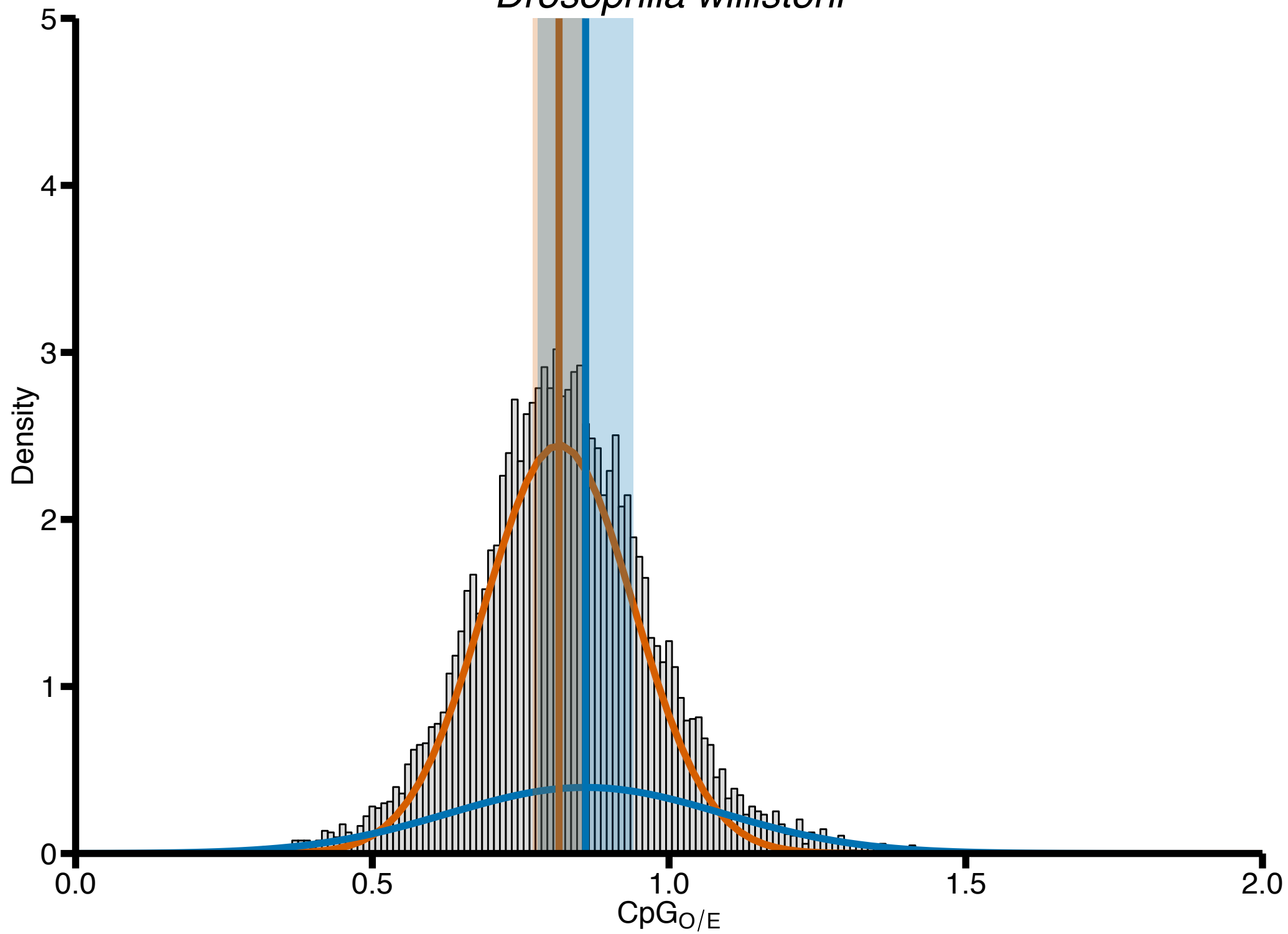
Drosophila yakuba
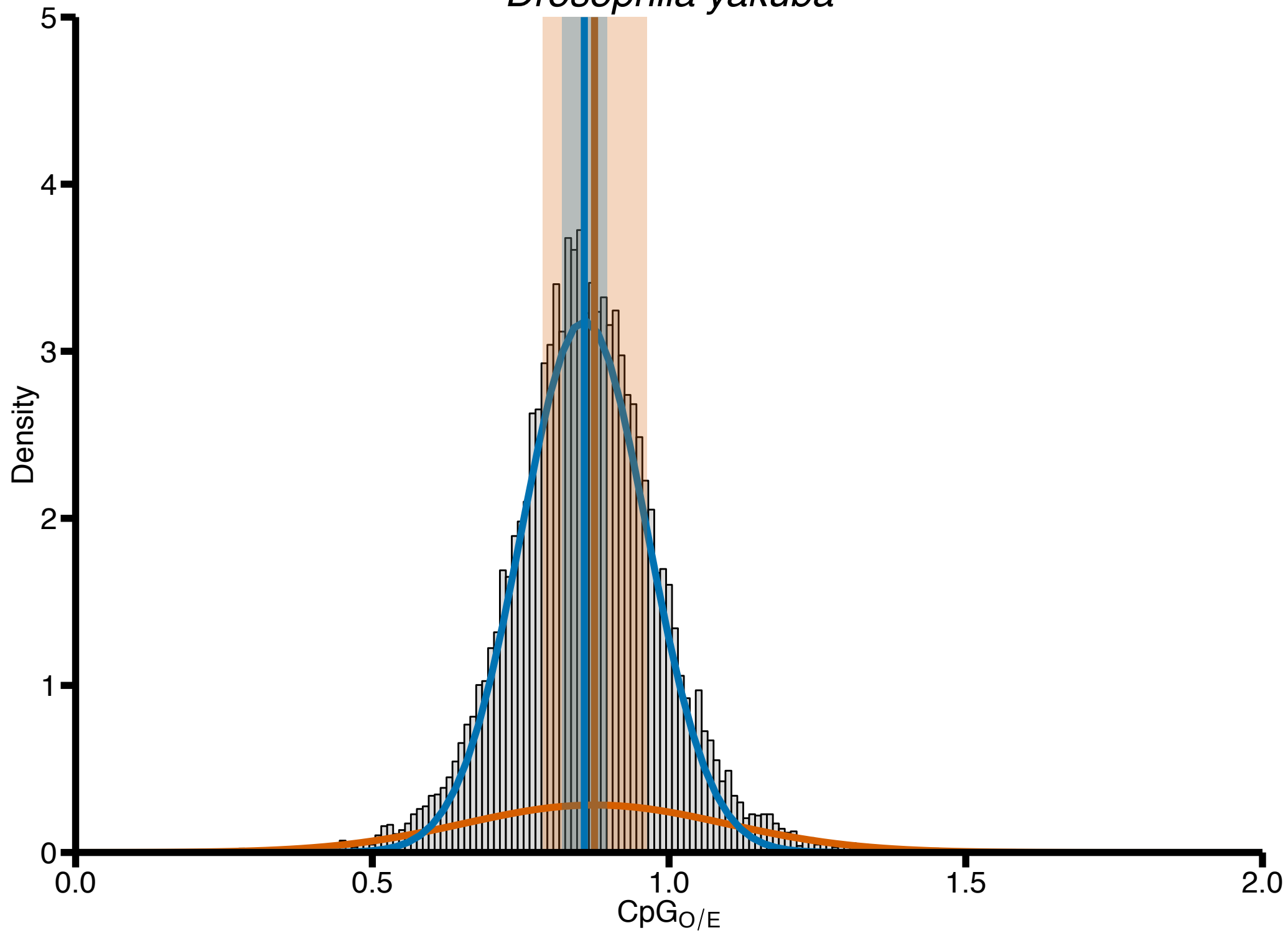
Glossina austeni
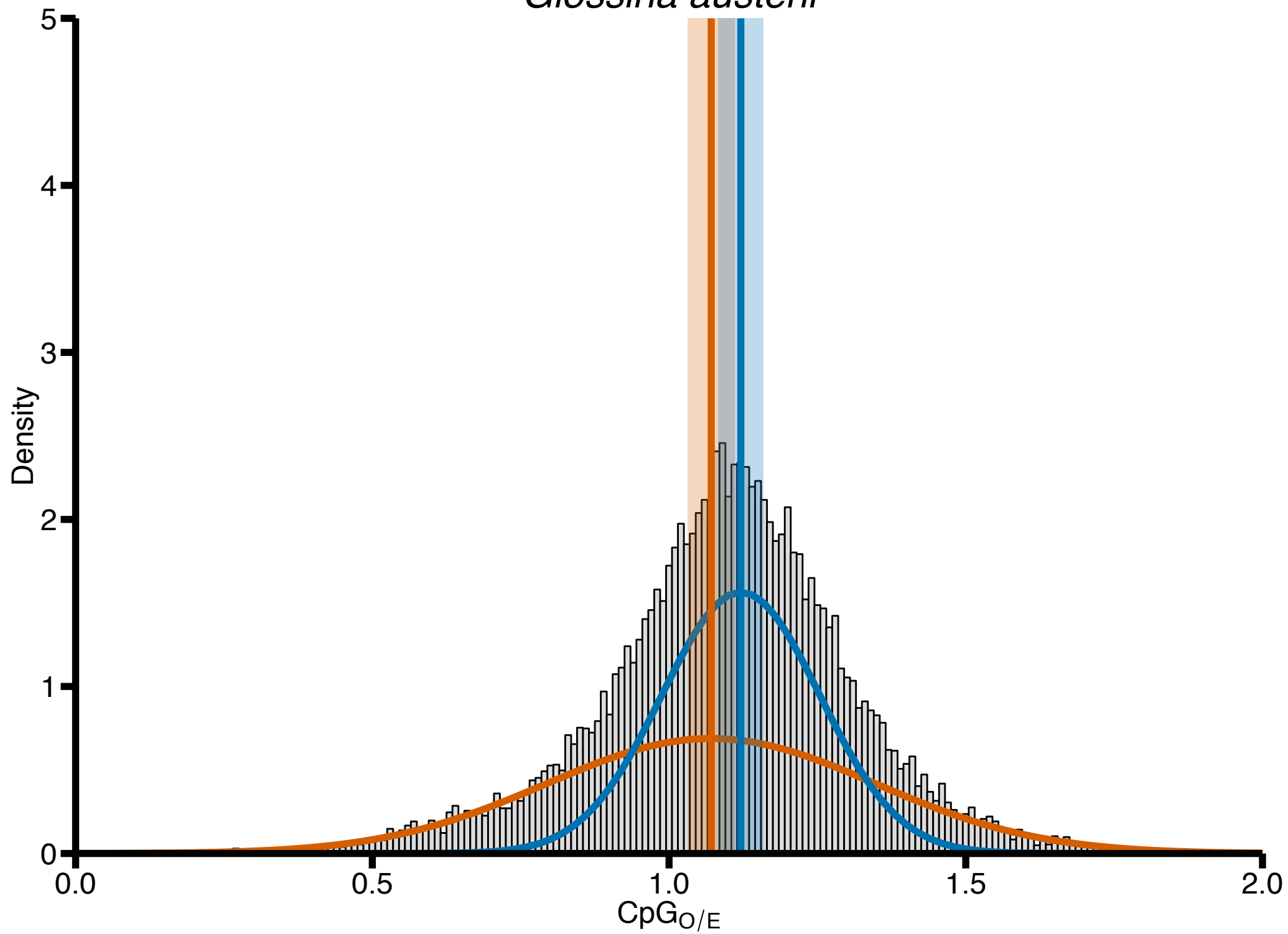
Glossina brevipalpis
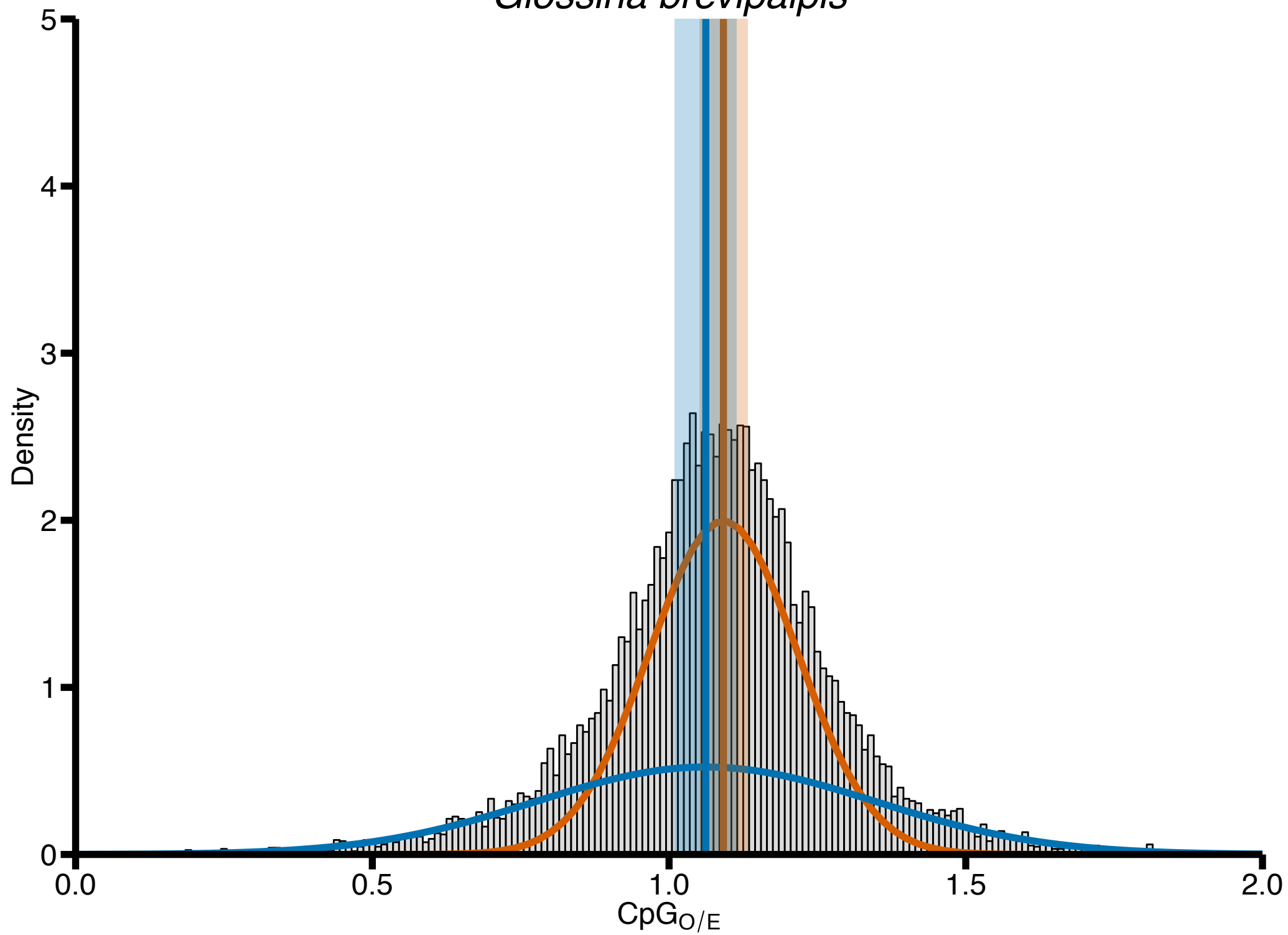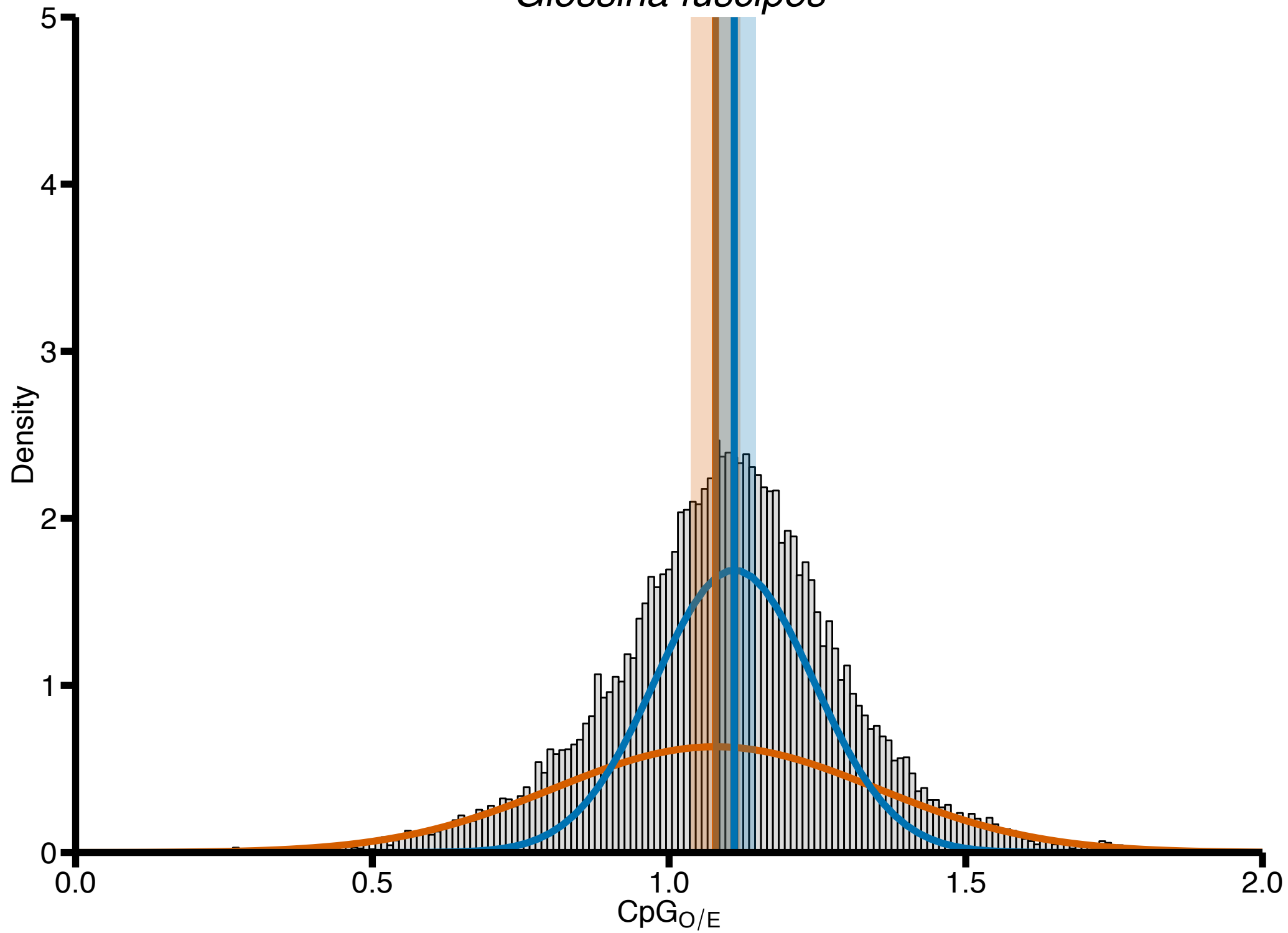
Glossina morsitans
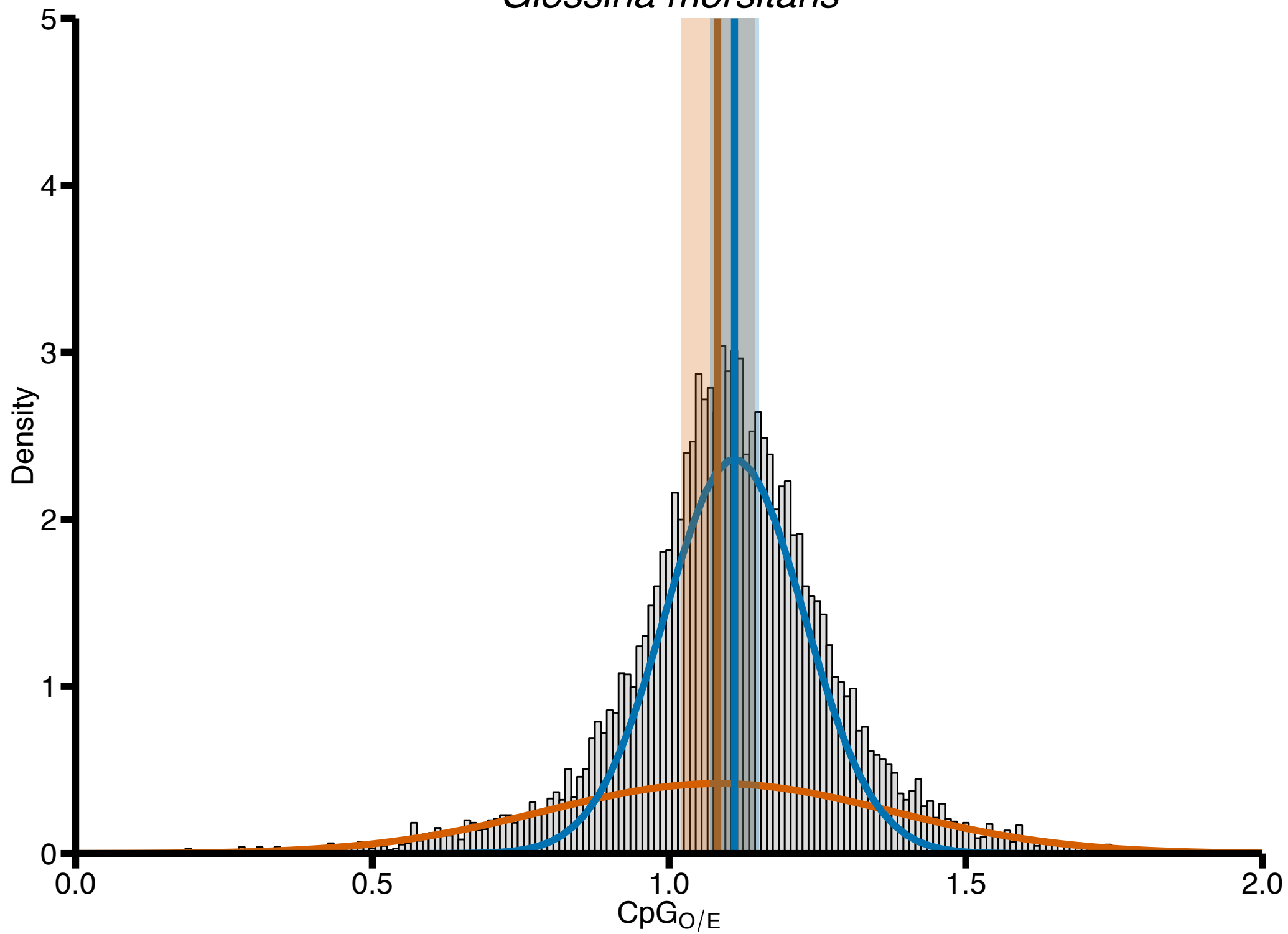
Glossina pallidipes
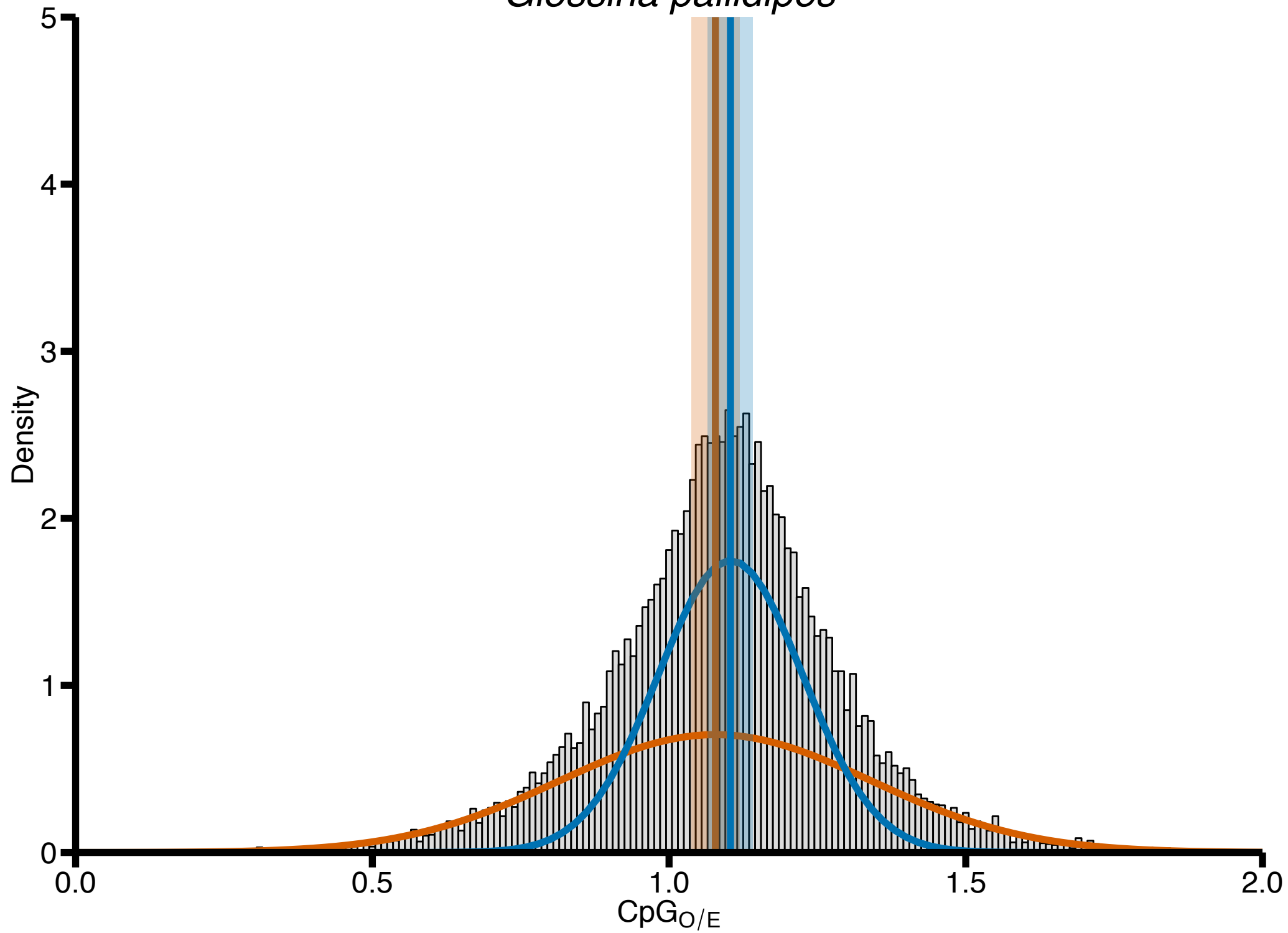
Glossina palpalis
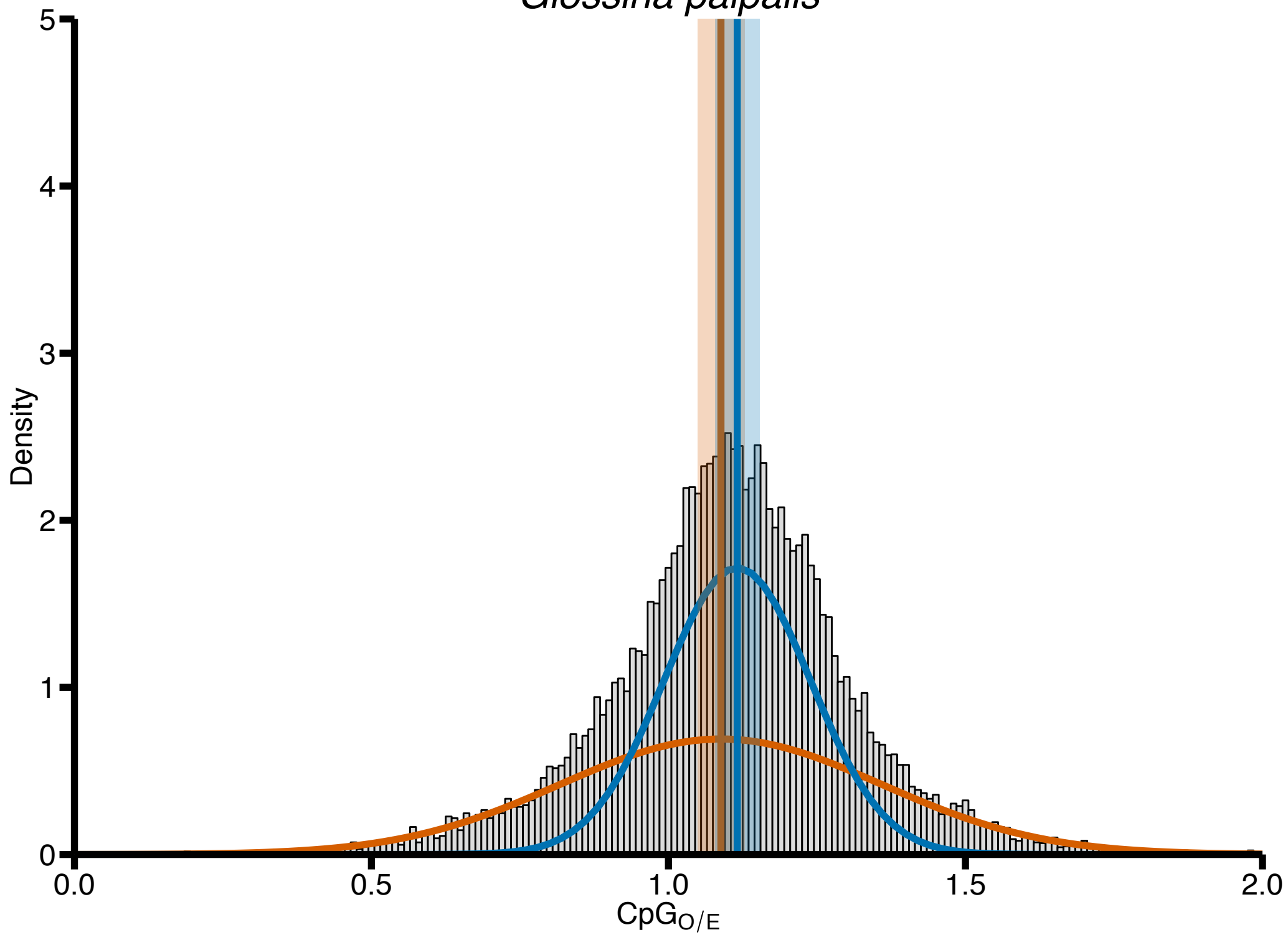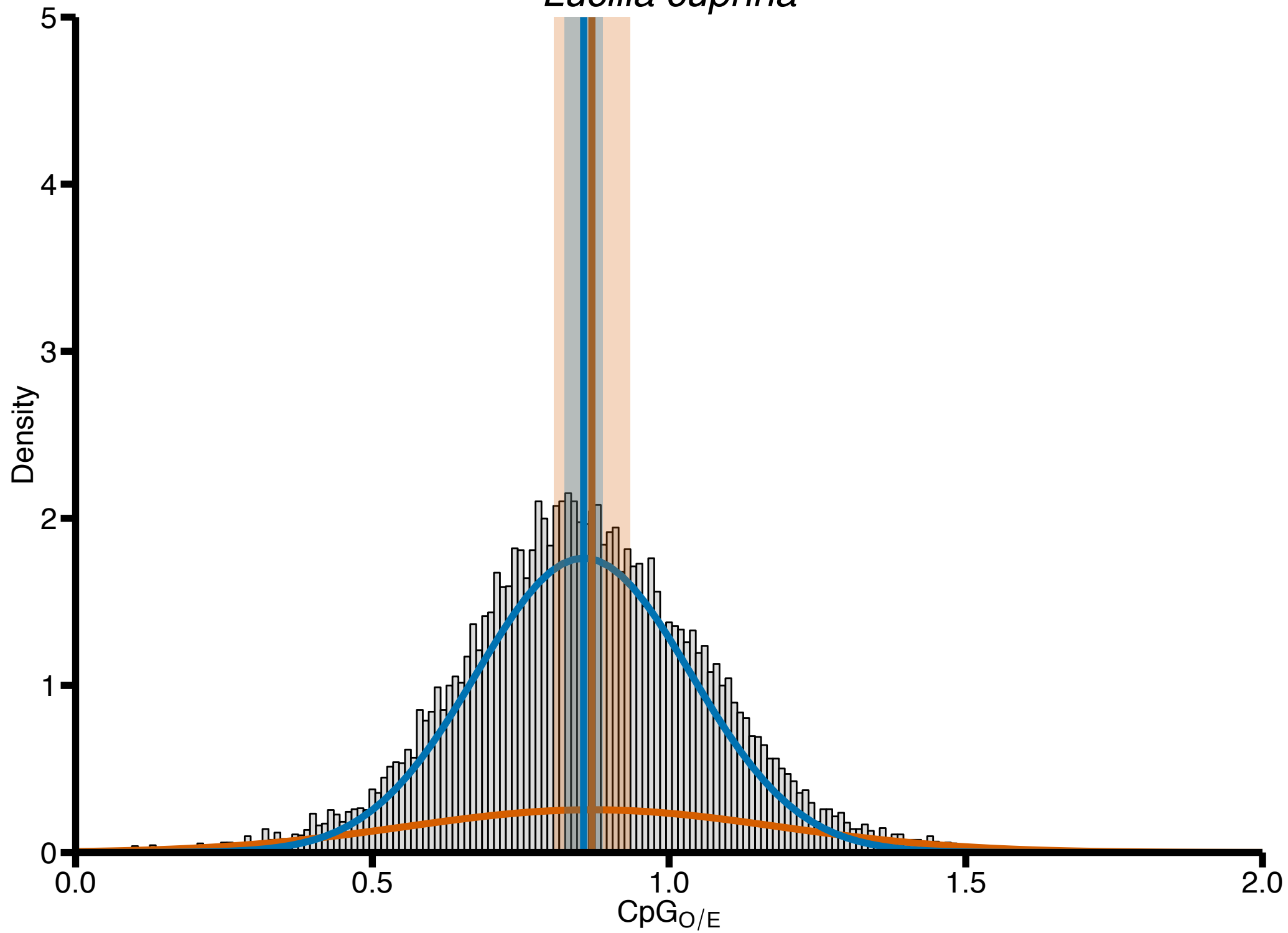
Lutzomyia longipalpis
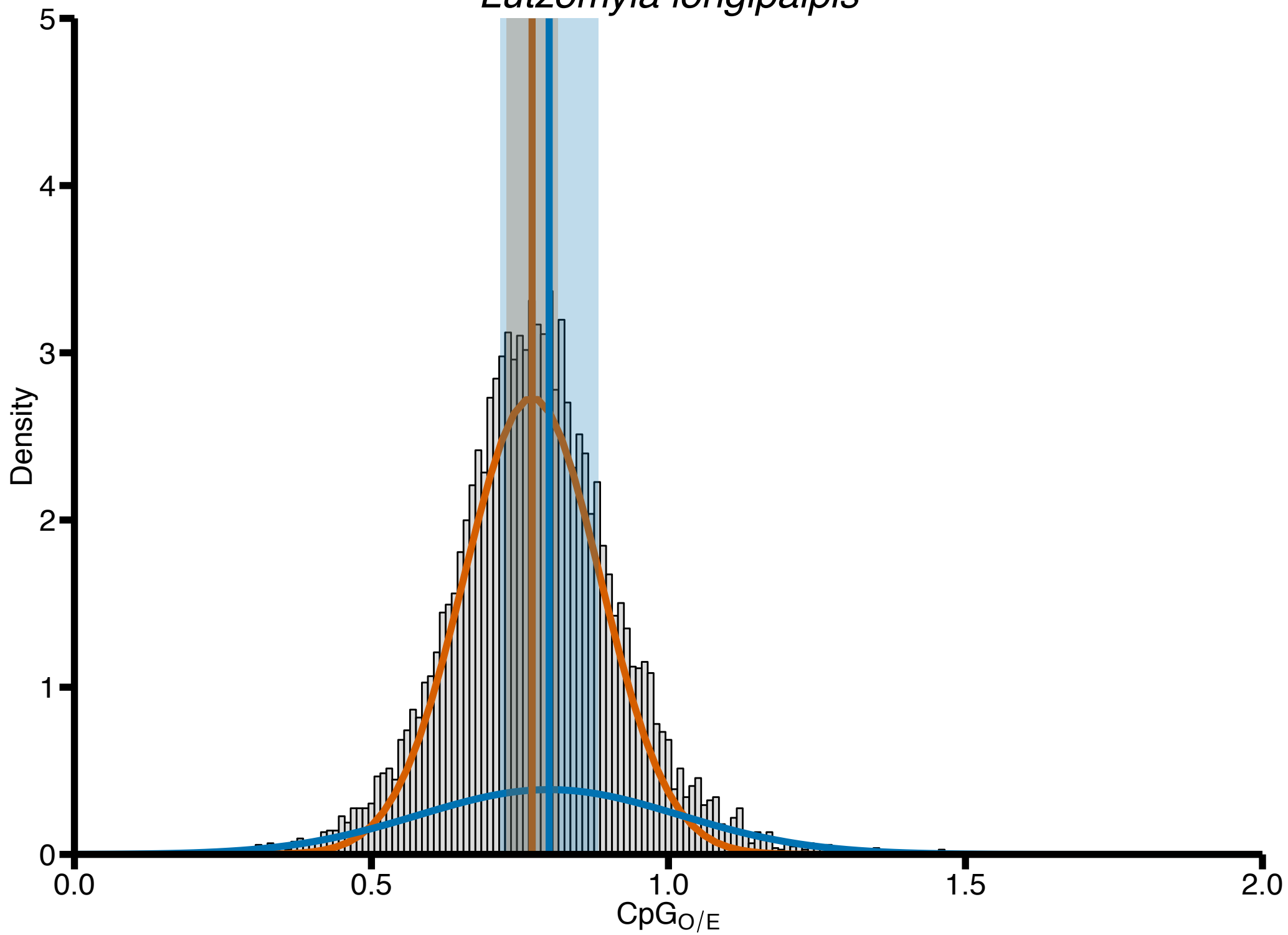
Musca domestica
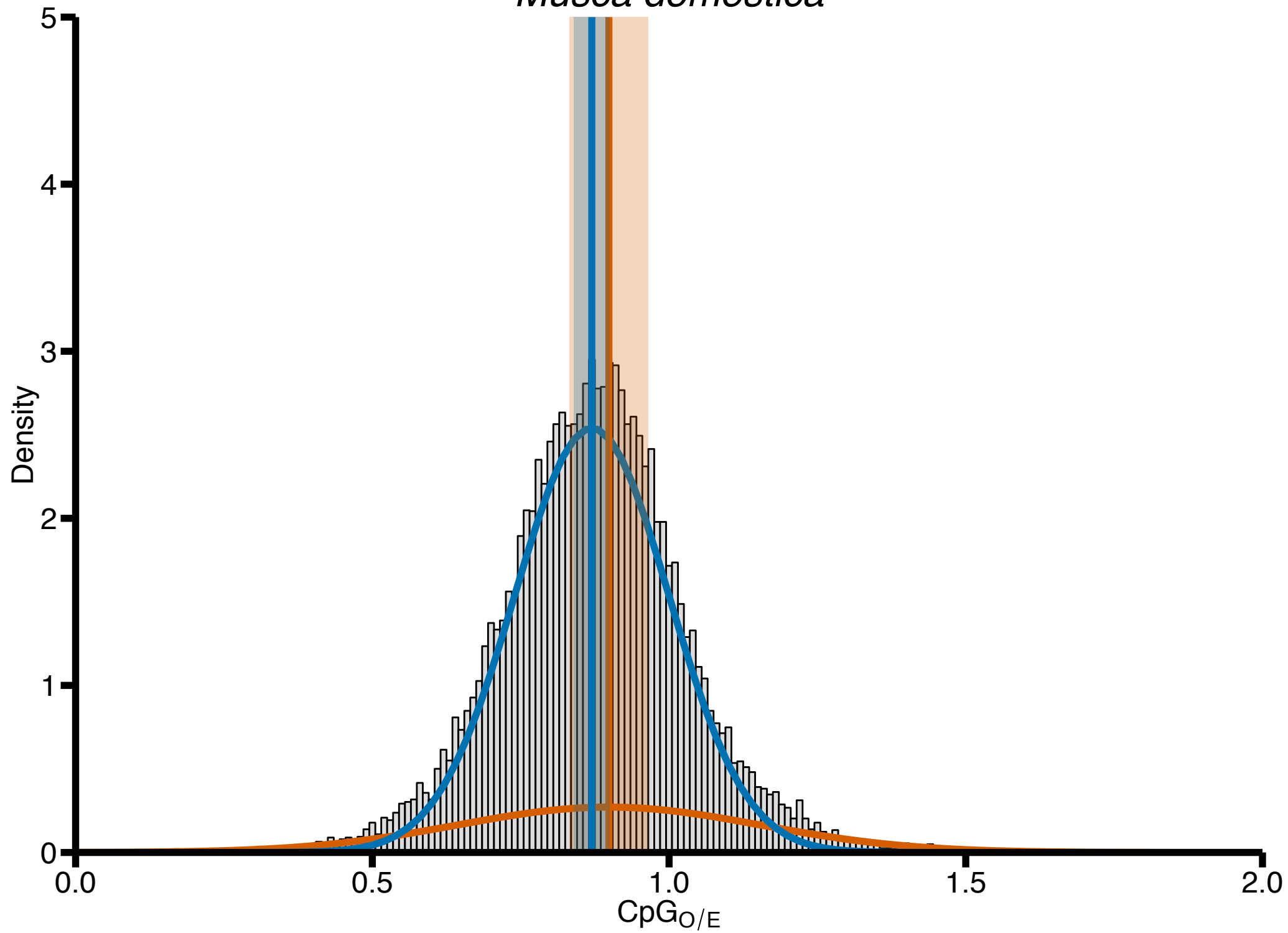
Phlebotomus papatasi
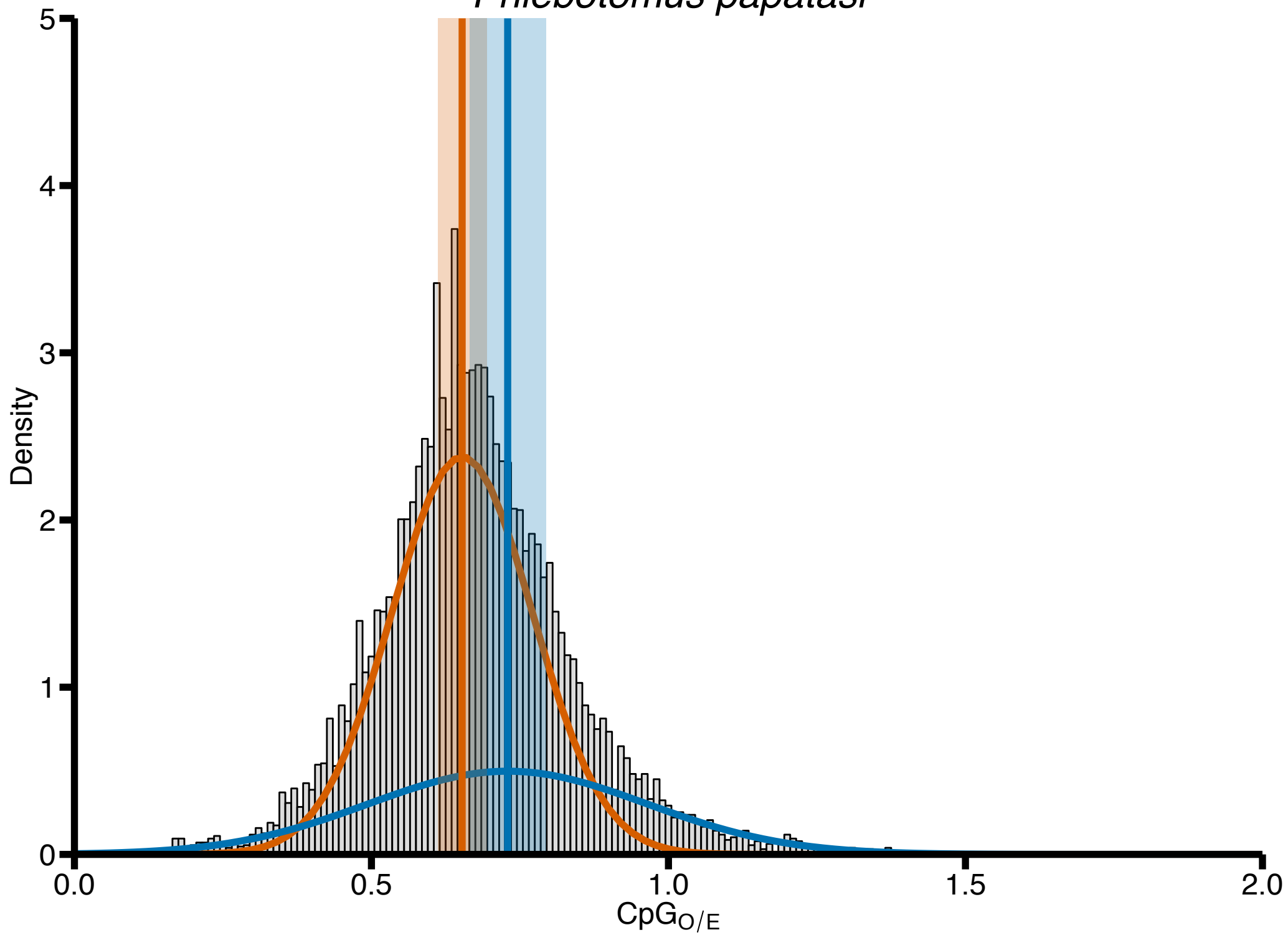
Stomoxys calcitrans
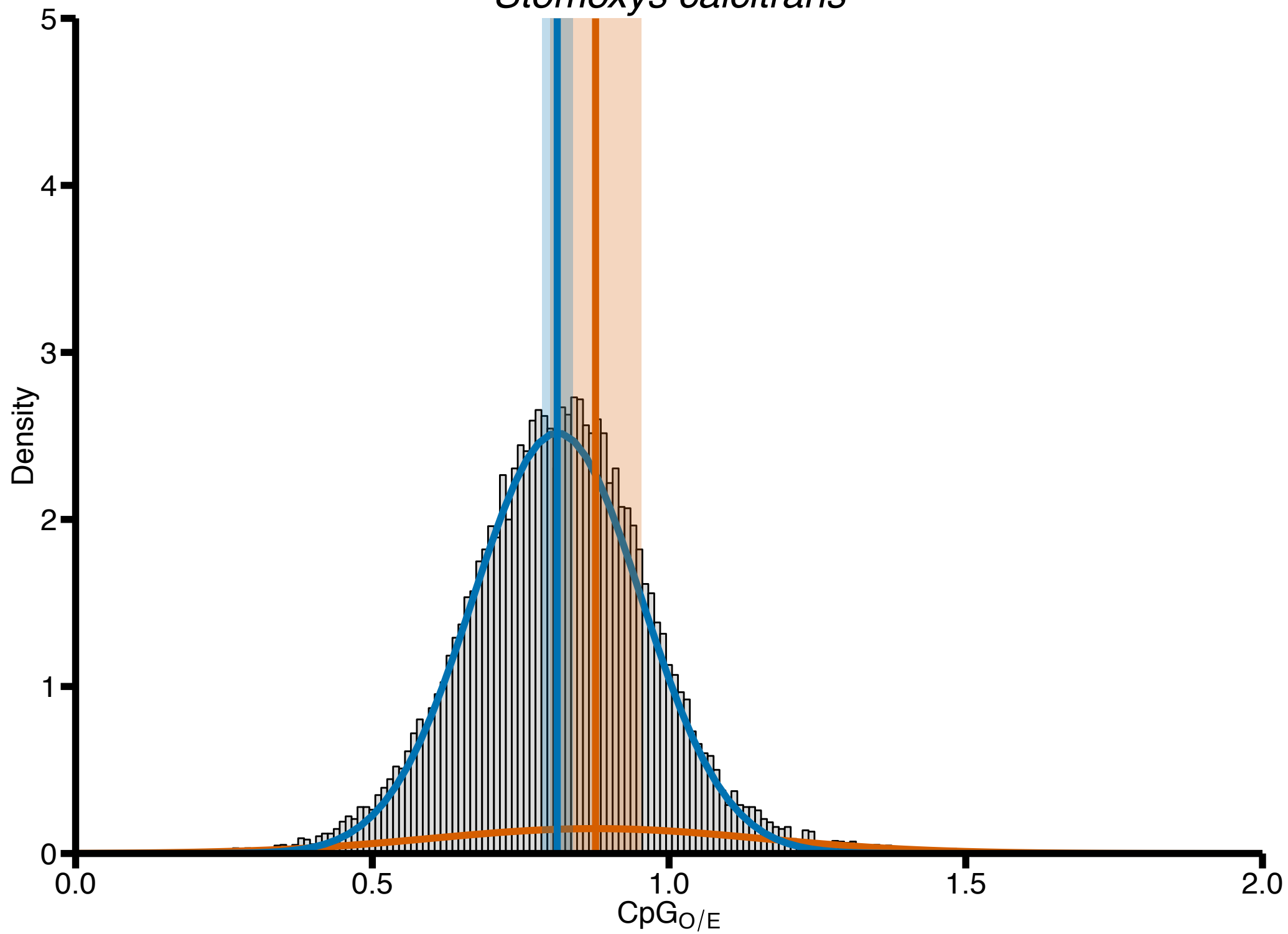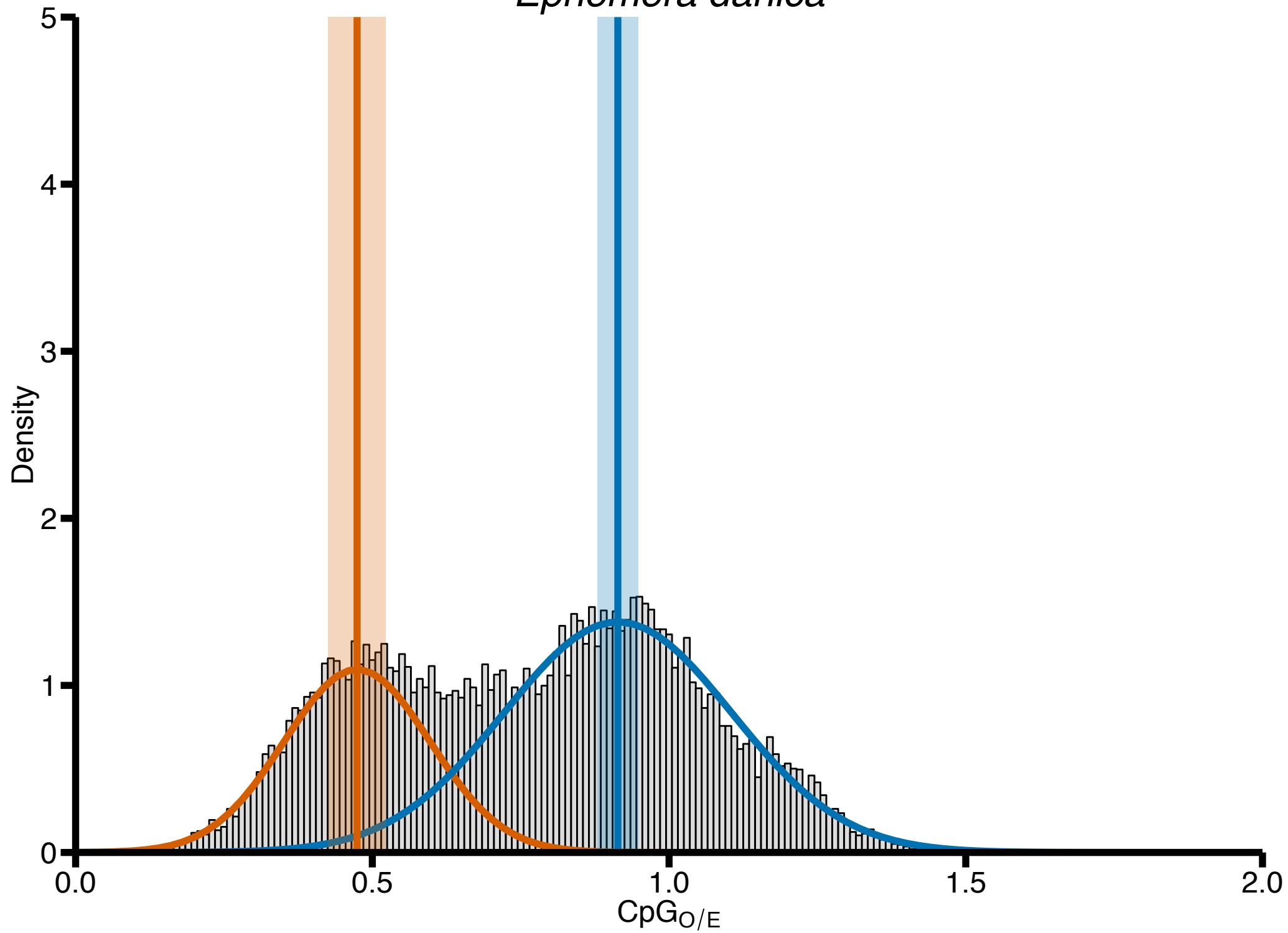
Acyrthosiphon pisum
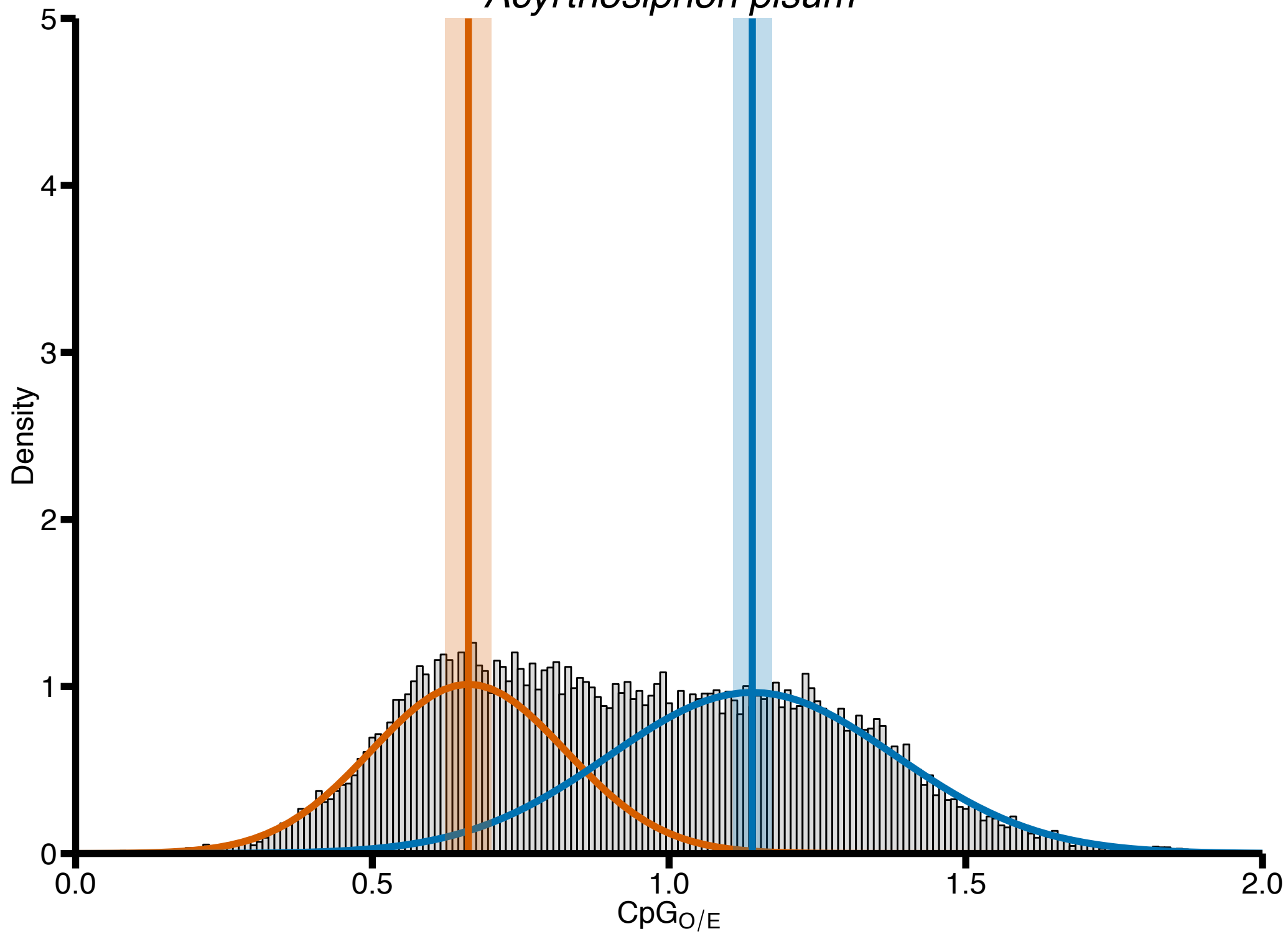
Cimex lectularius
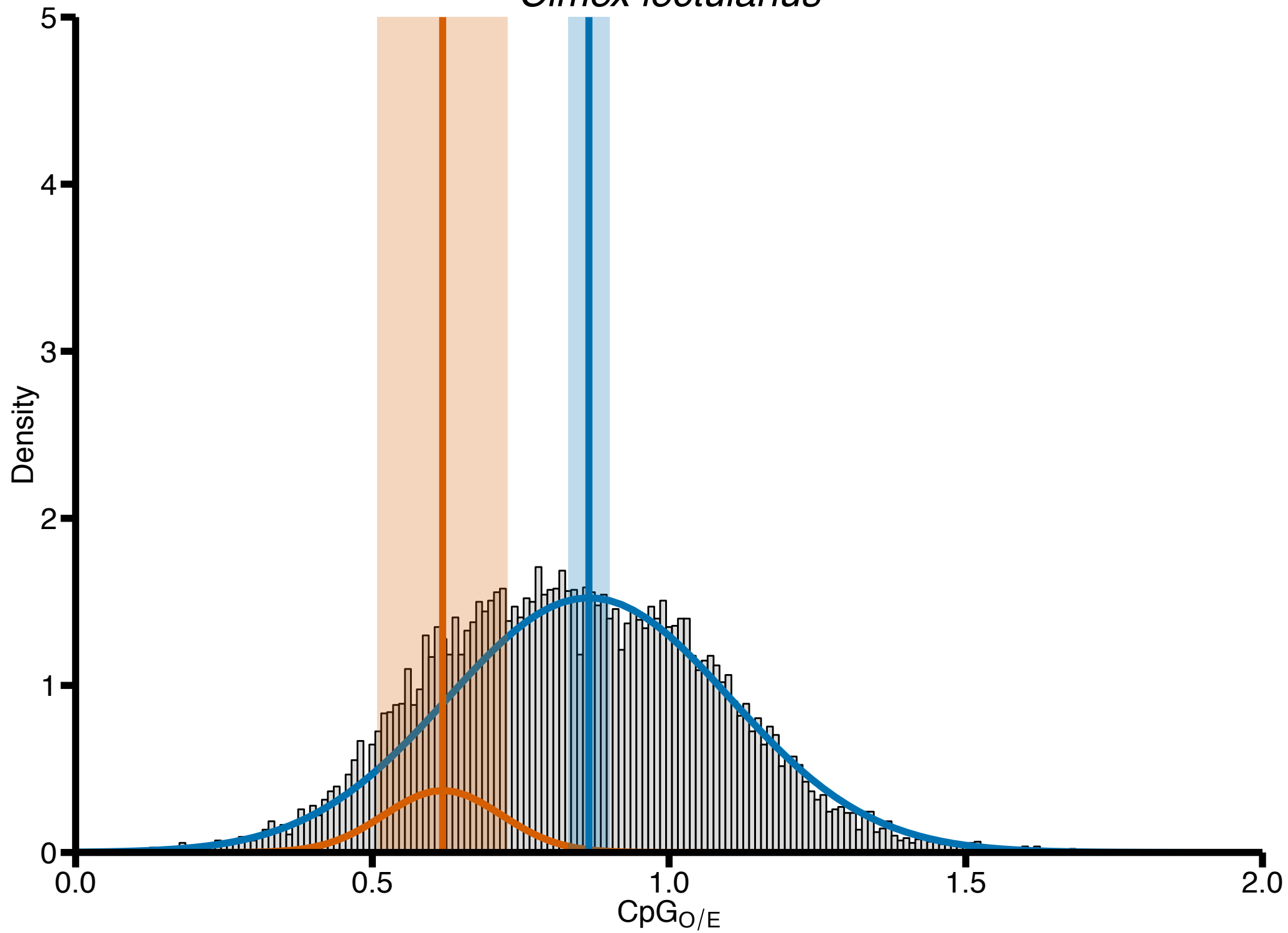
Diaphorina citri
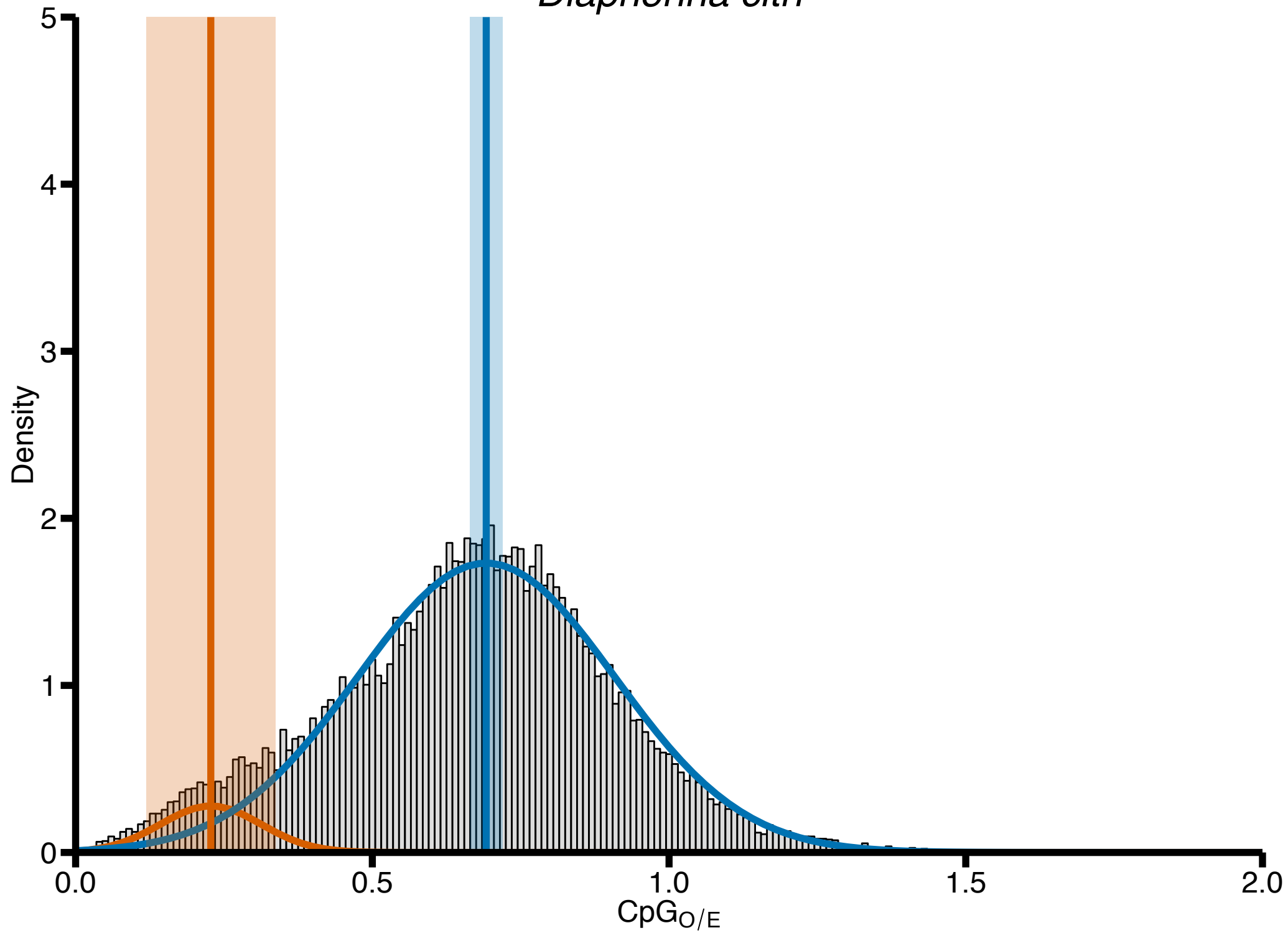
Gerris buenoi
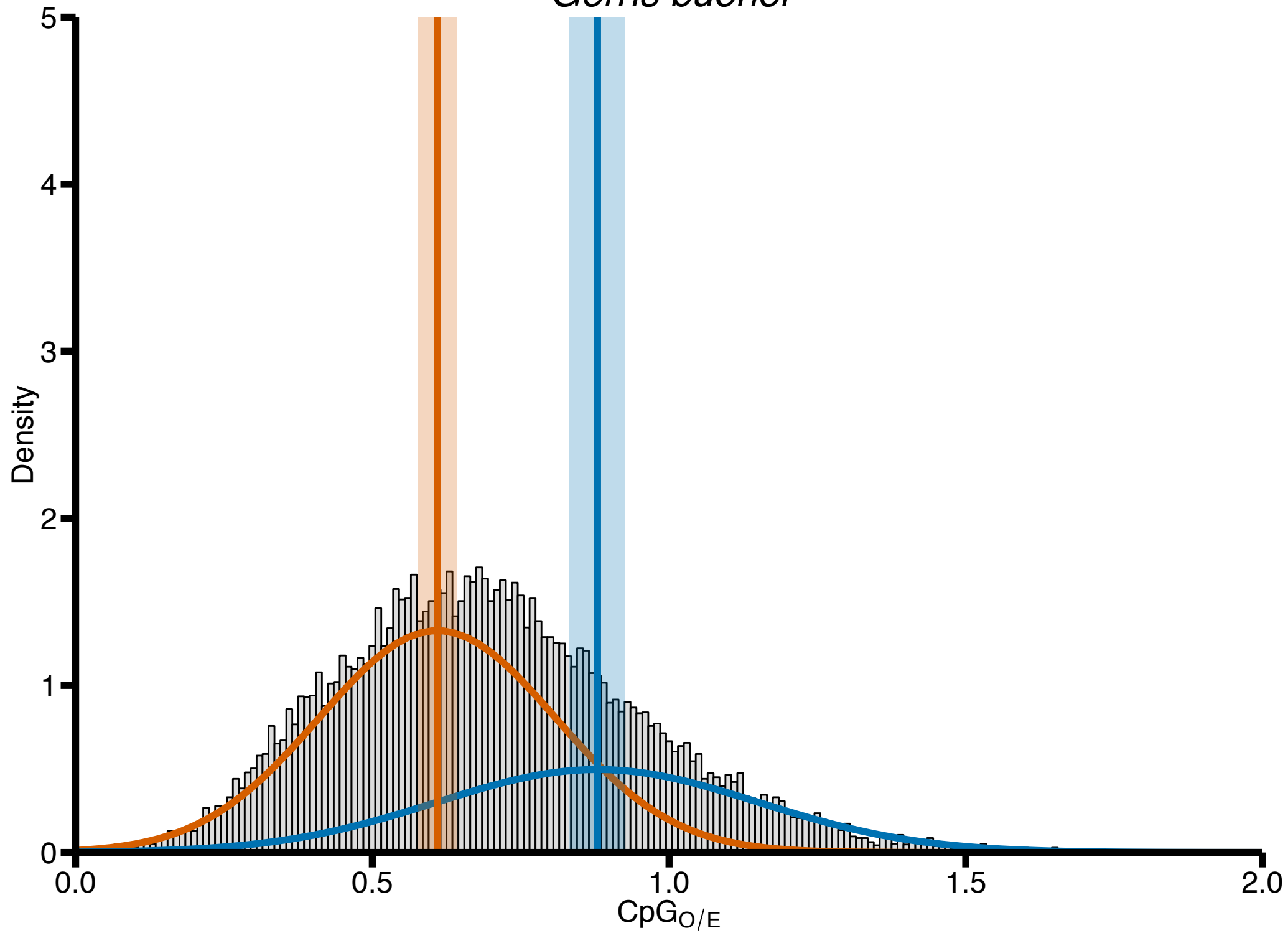
Halyomorpha halys
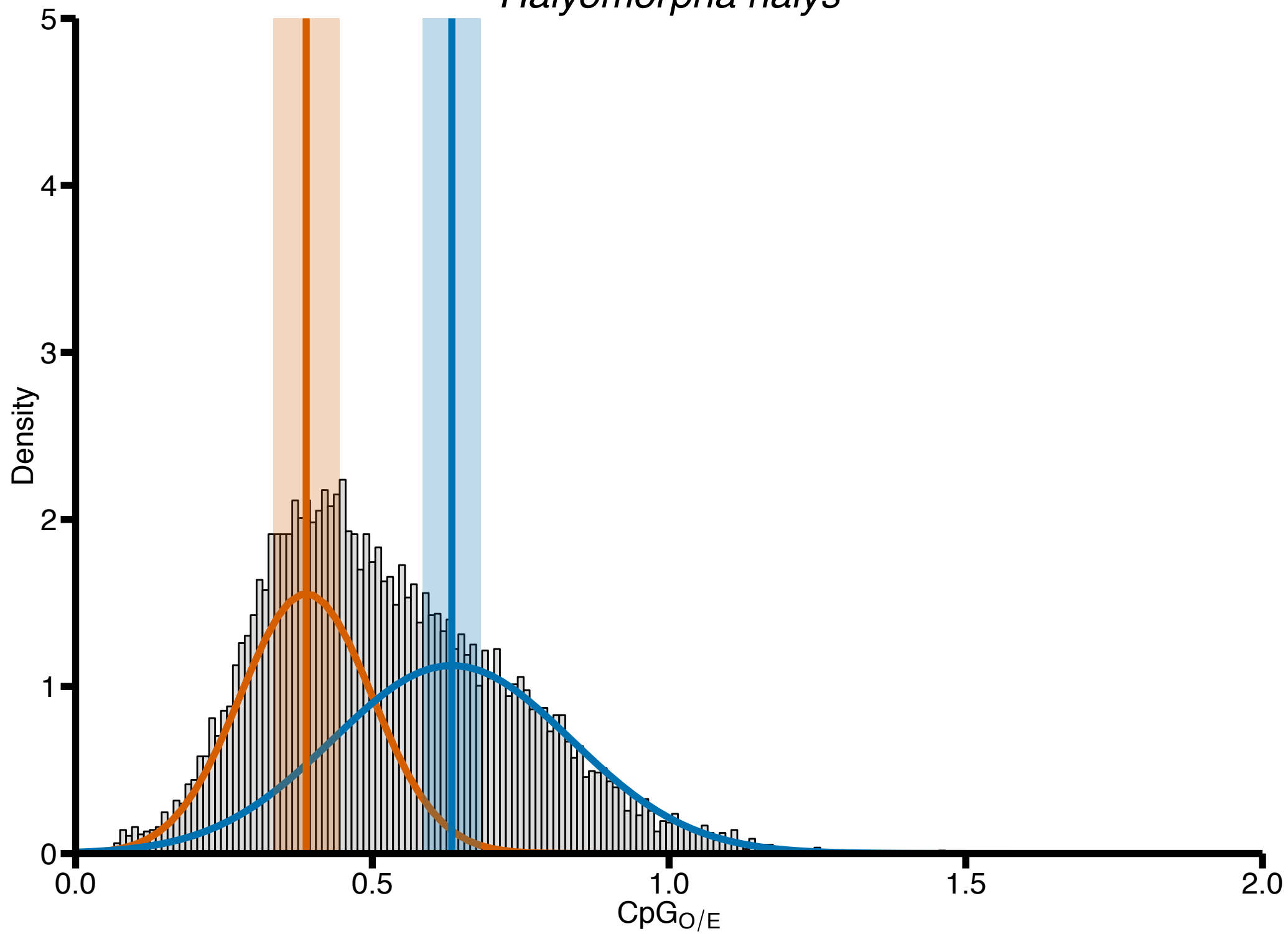
Homalodisca vitripennis
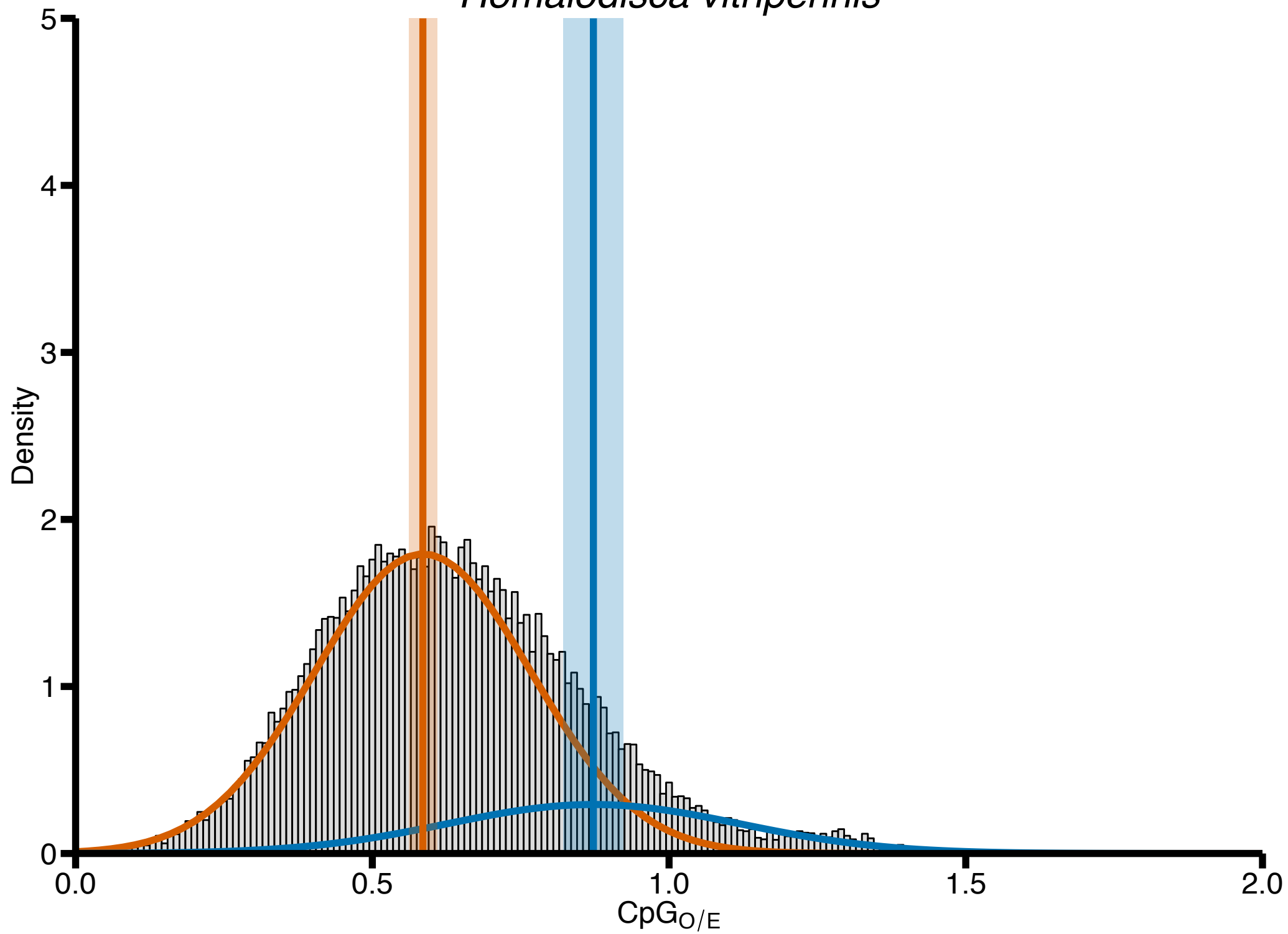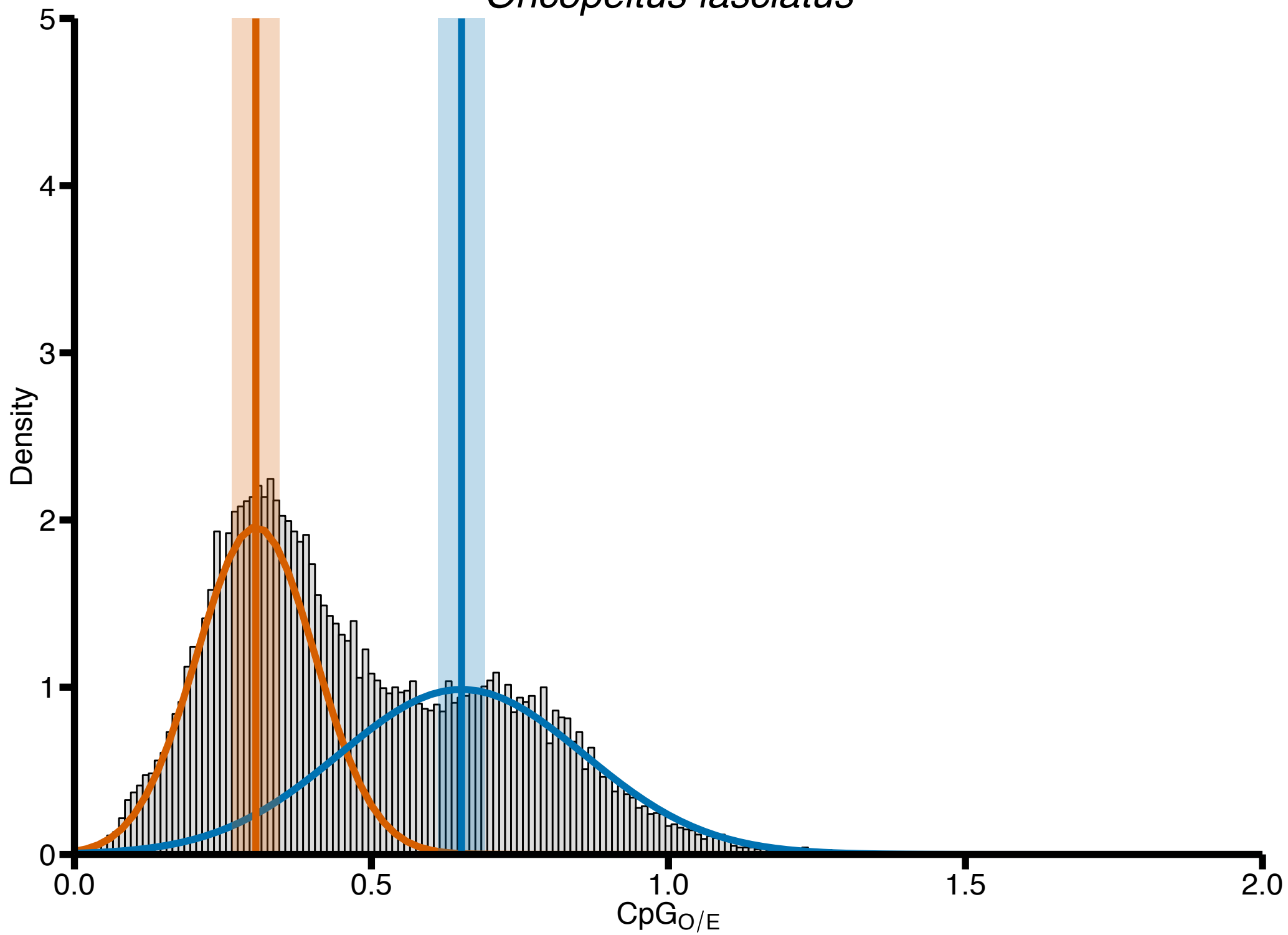
Pachypsylla venusta
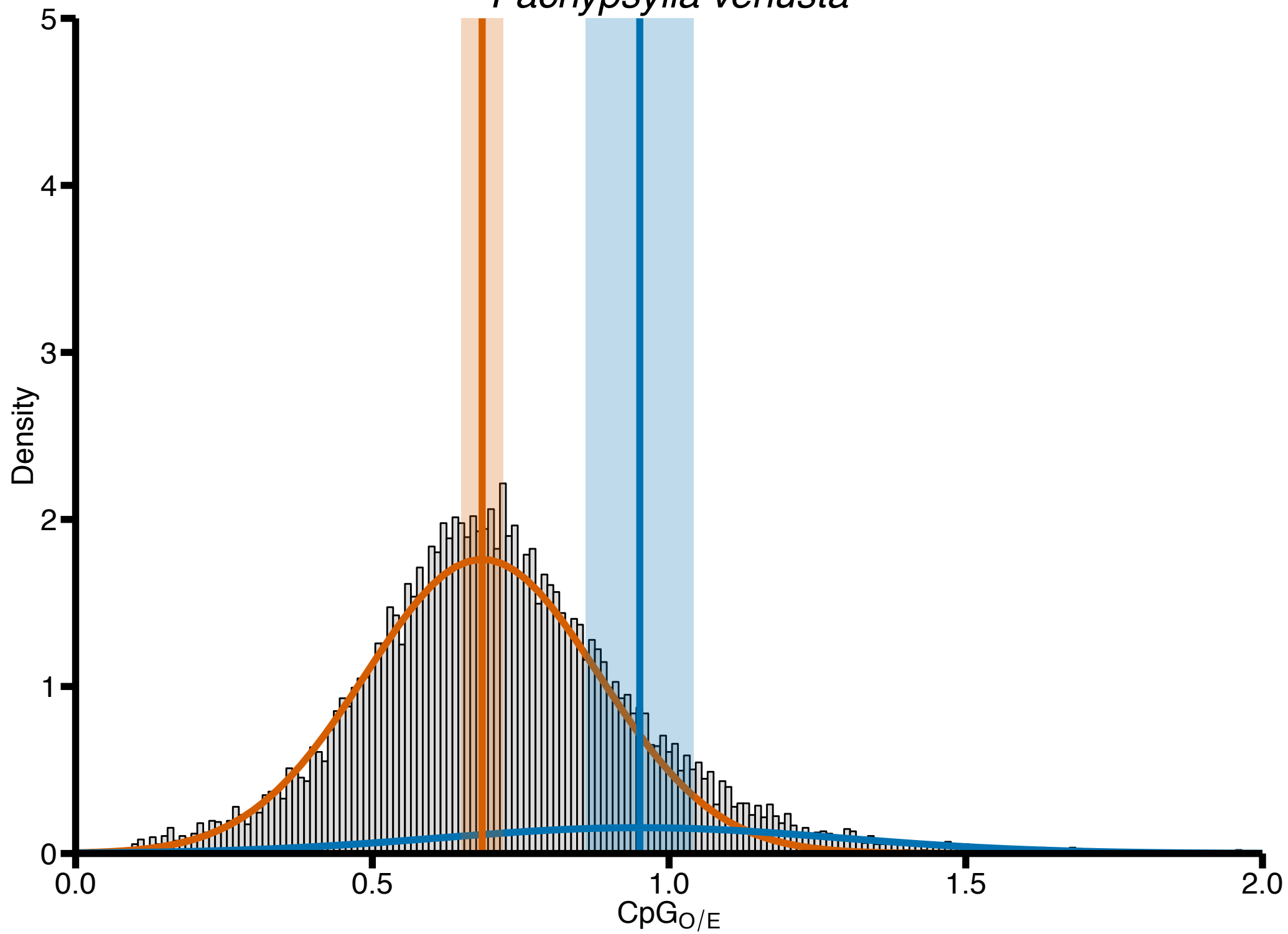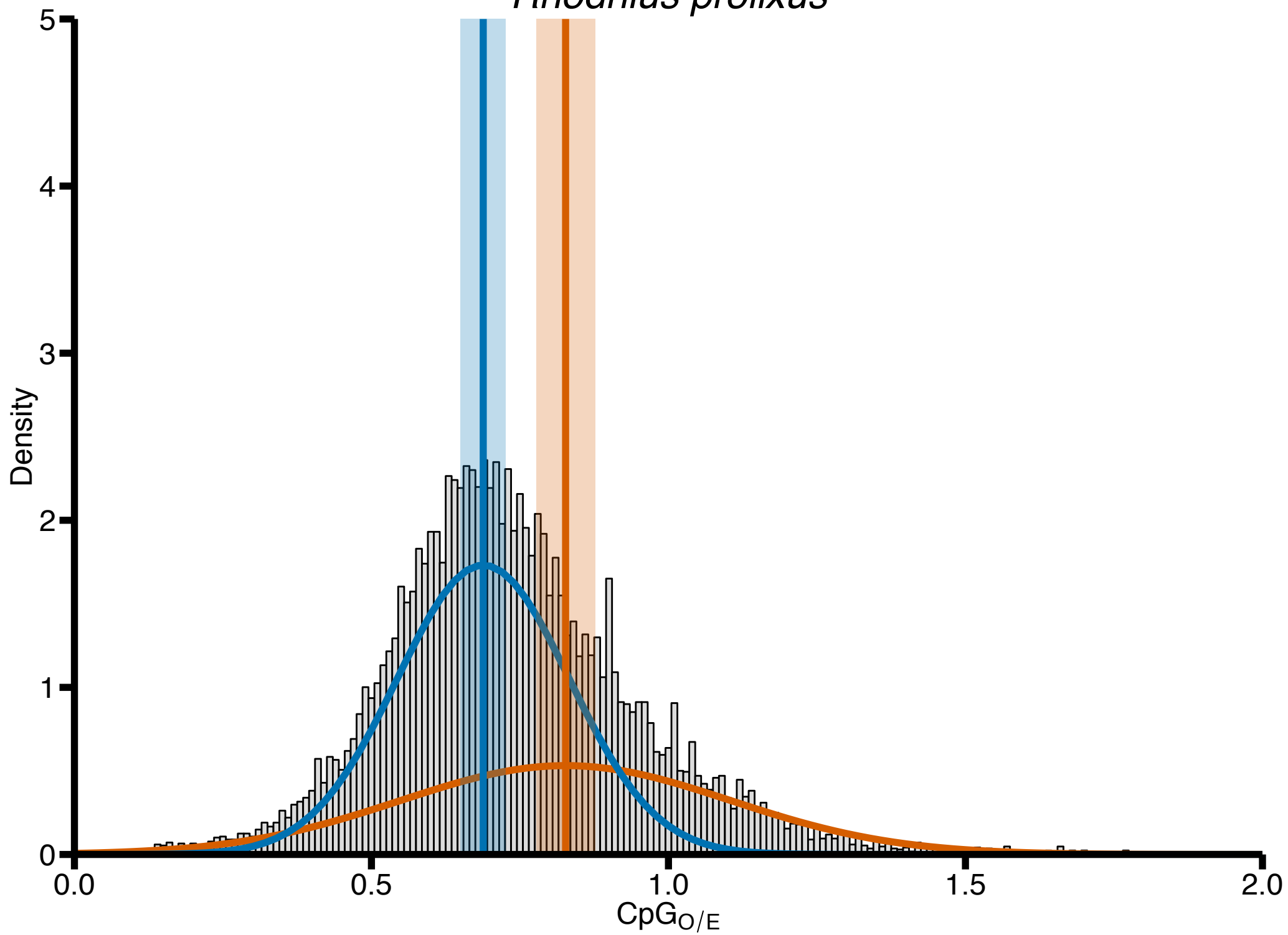
Acromyrmex echinatior
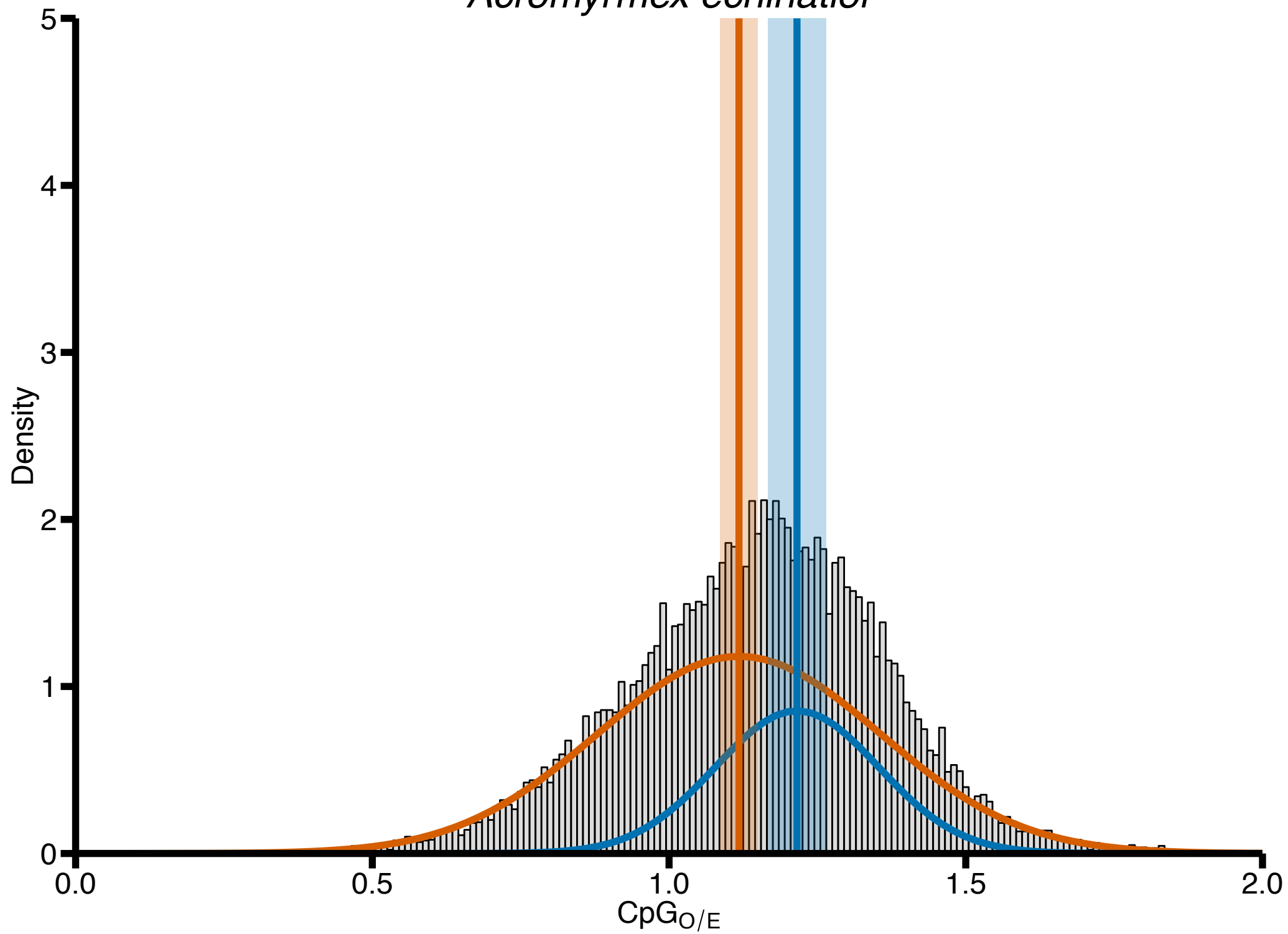
Apis dorsata
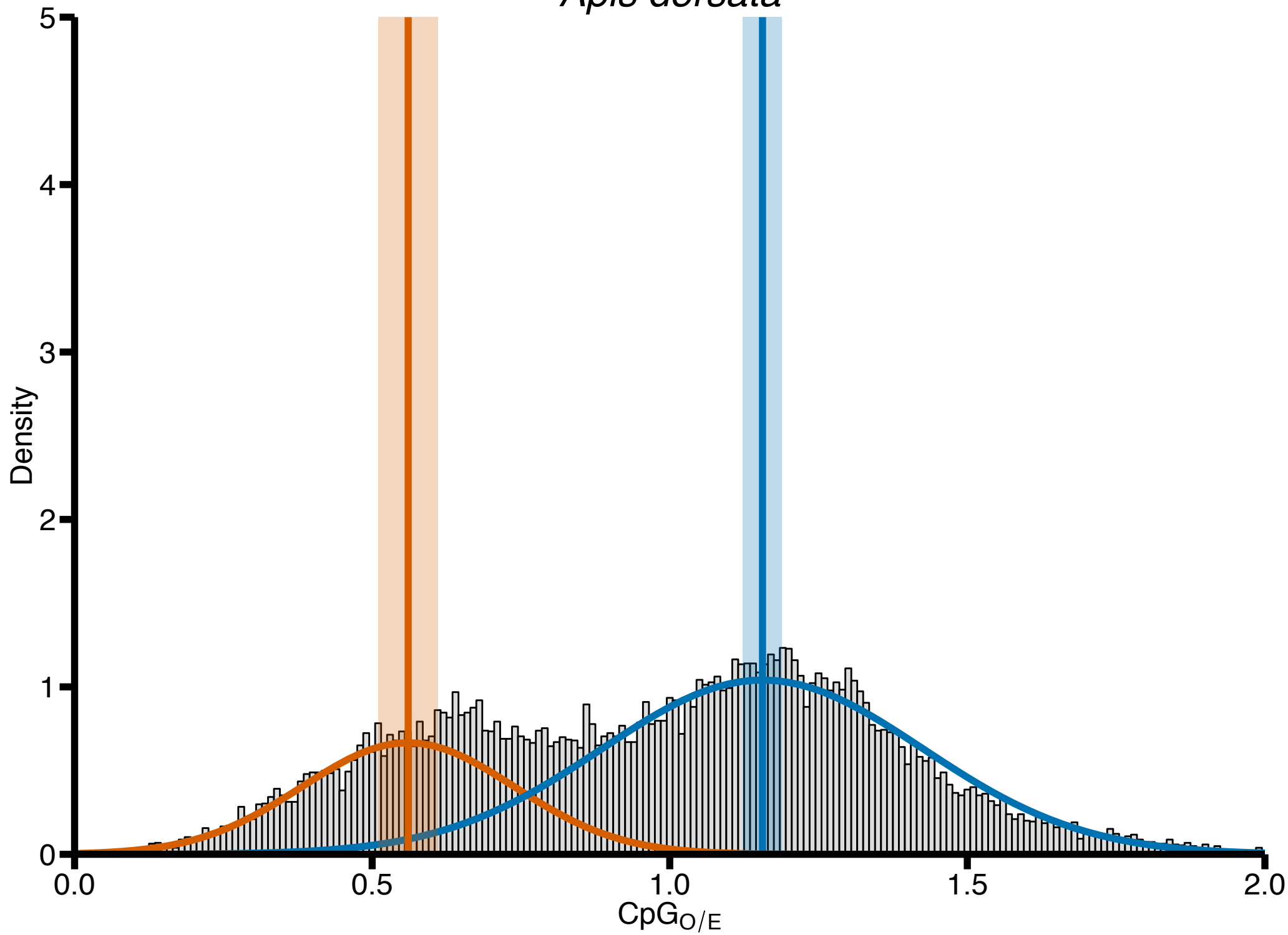
Apis florea
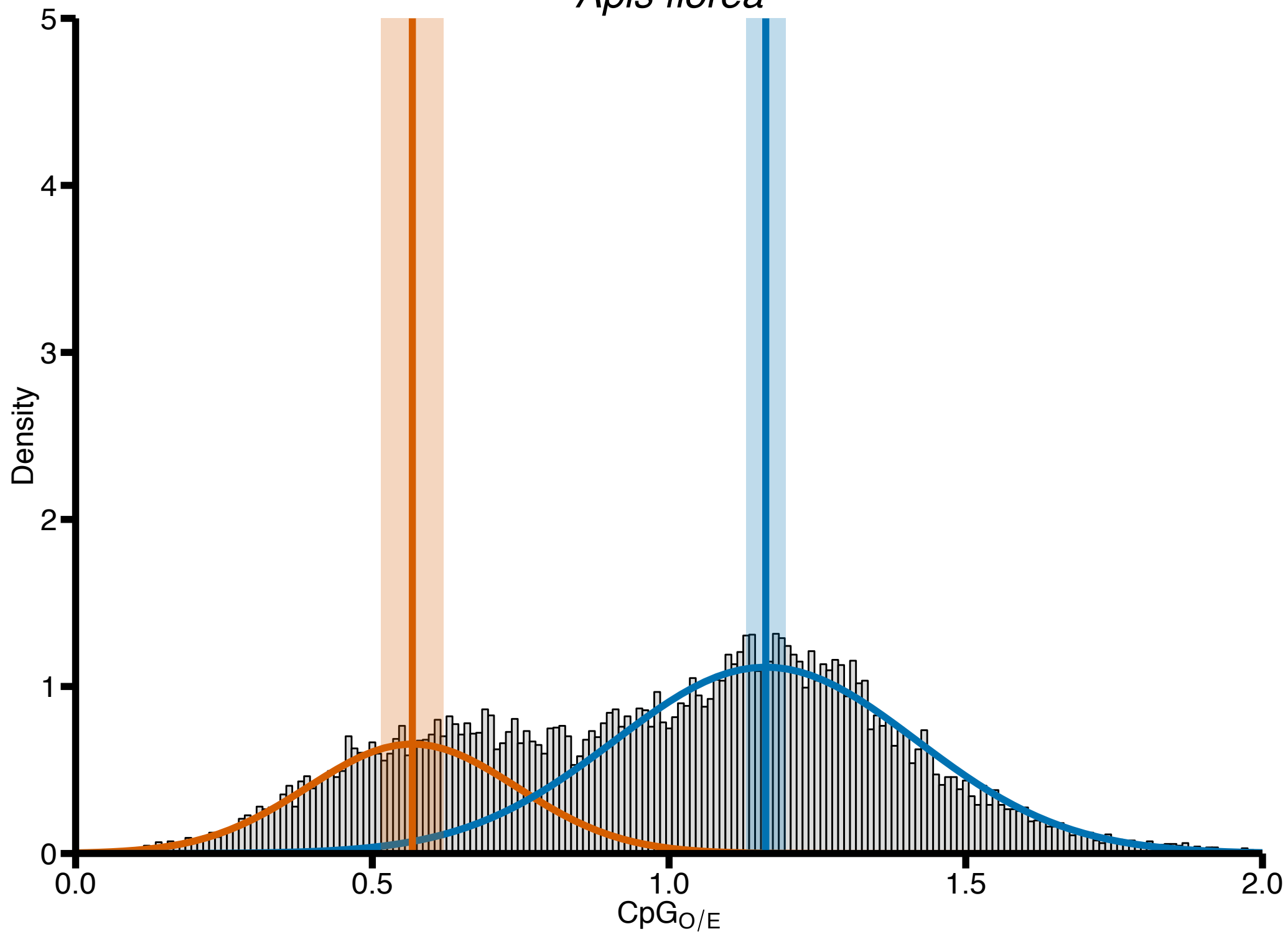
Apis mellifera
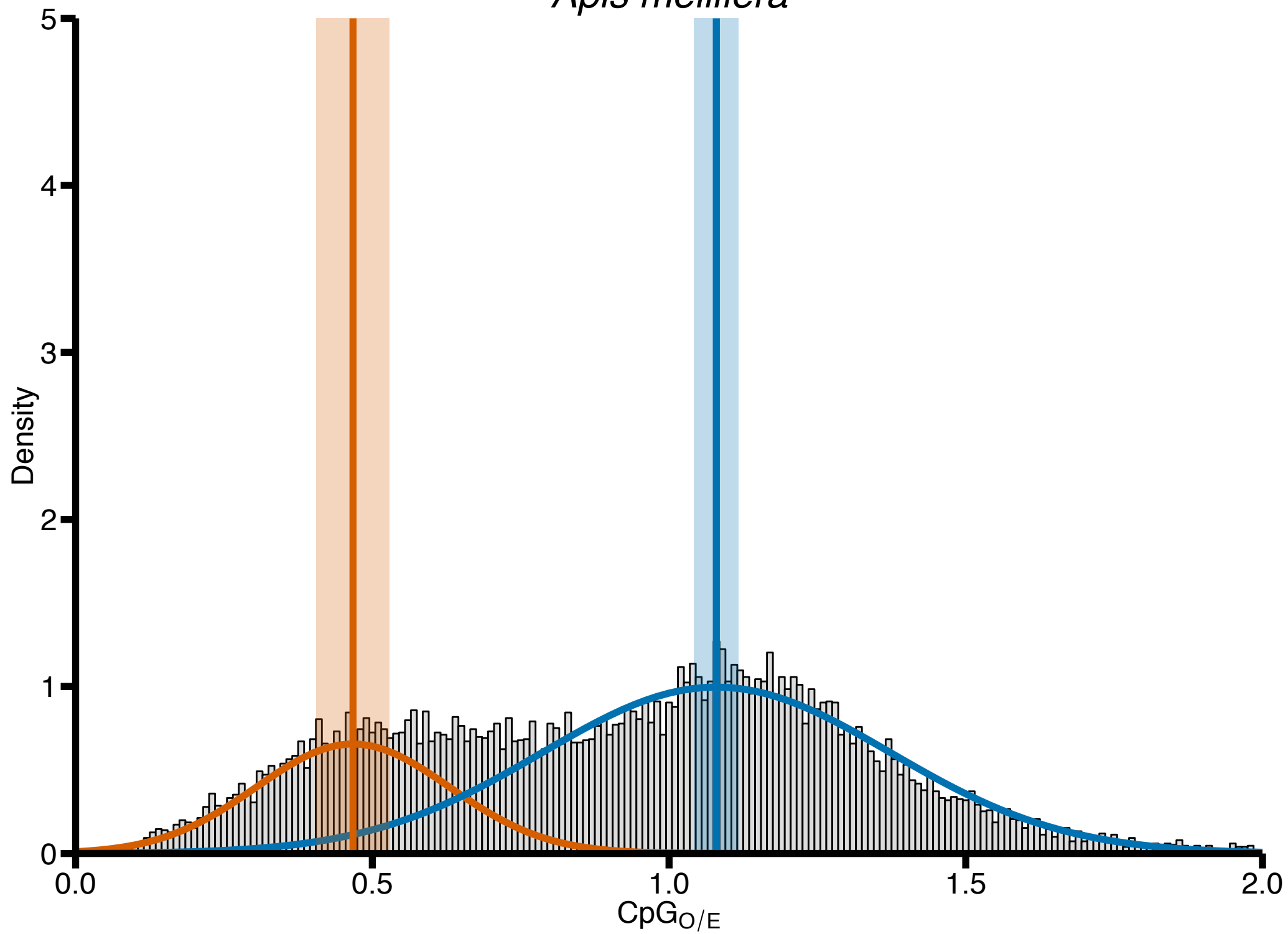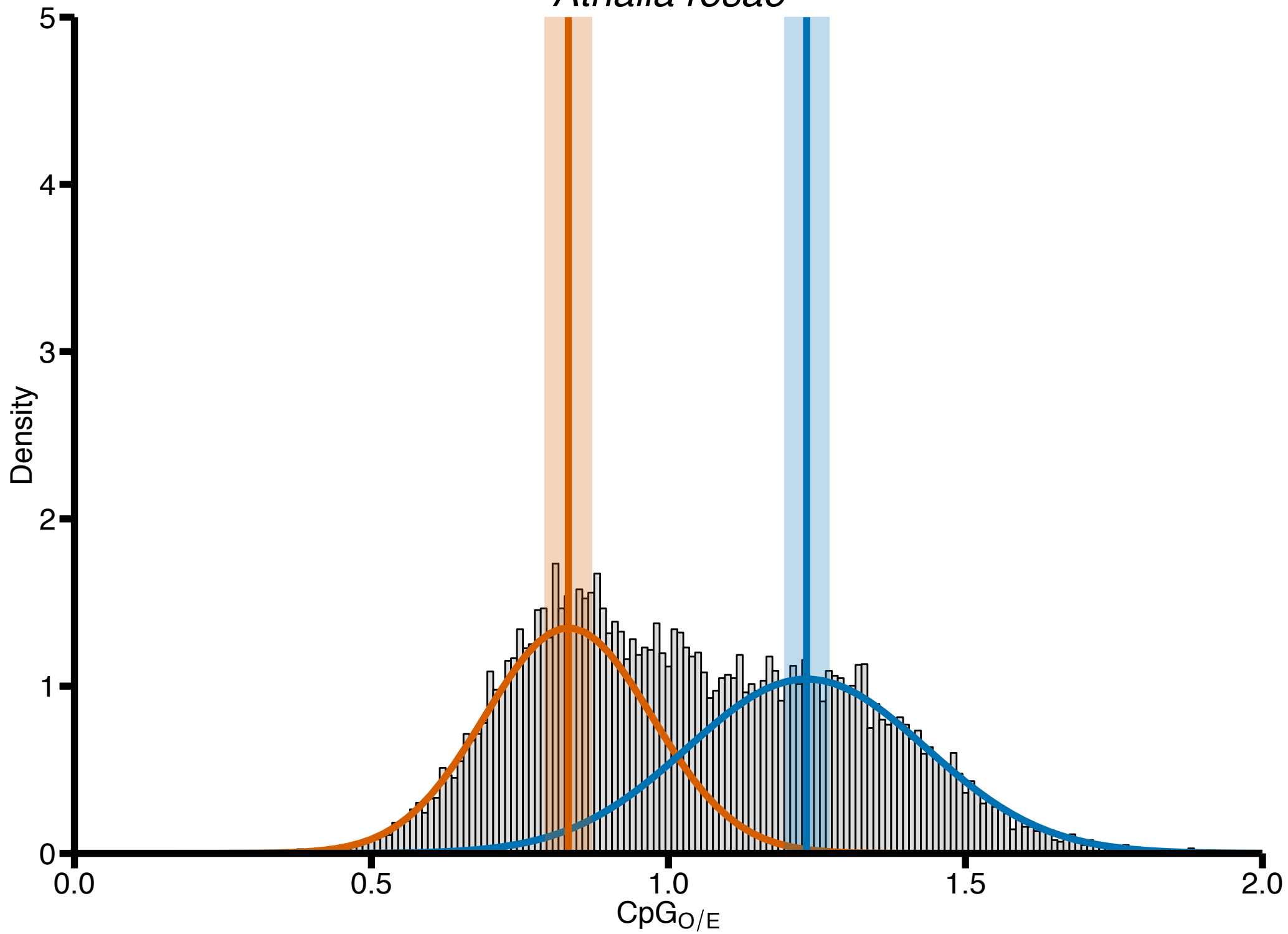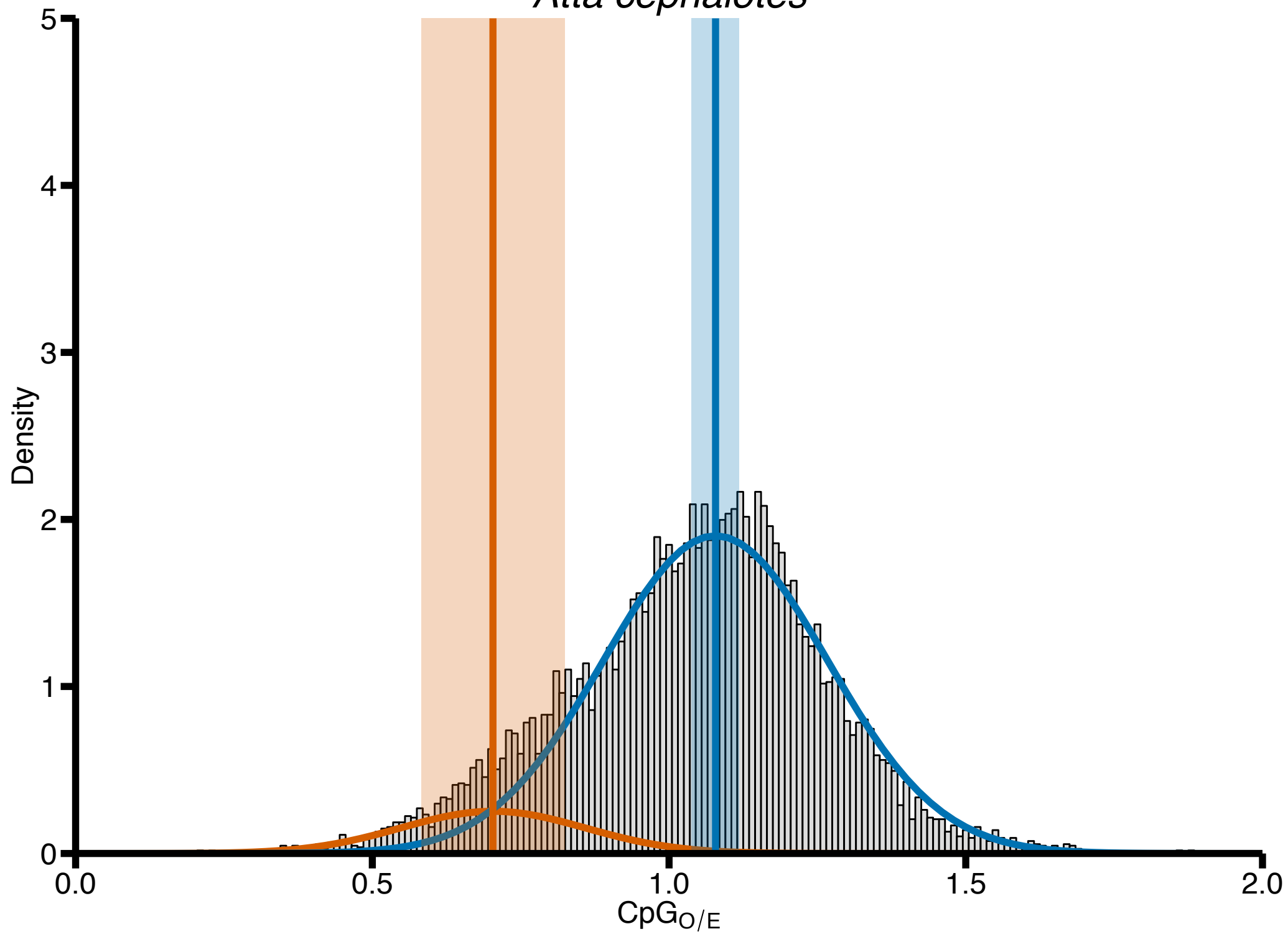
Bombus impatiens
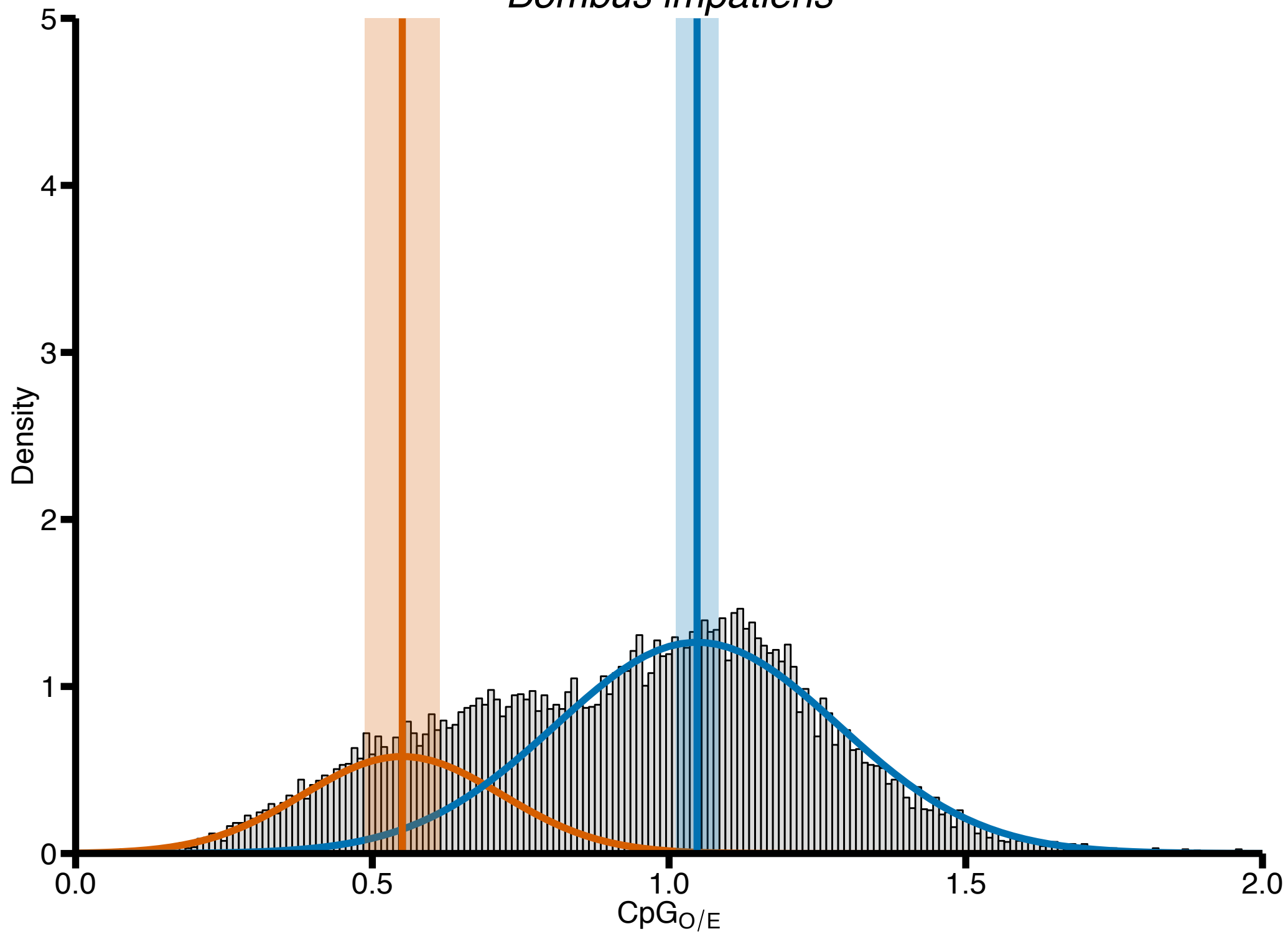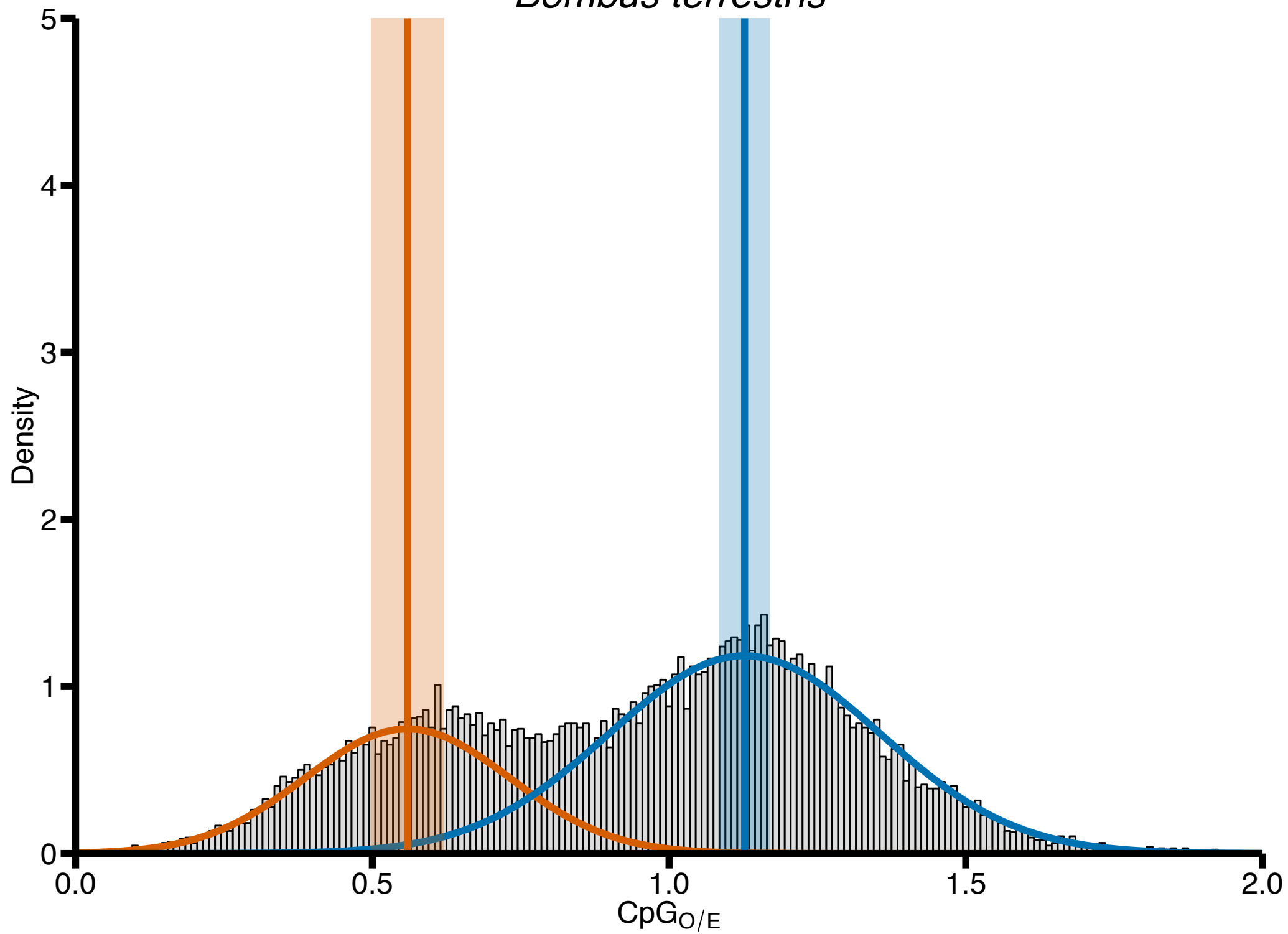
Camponotus floridanus
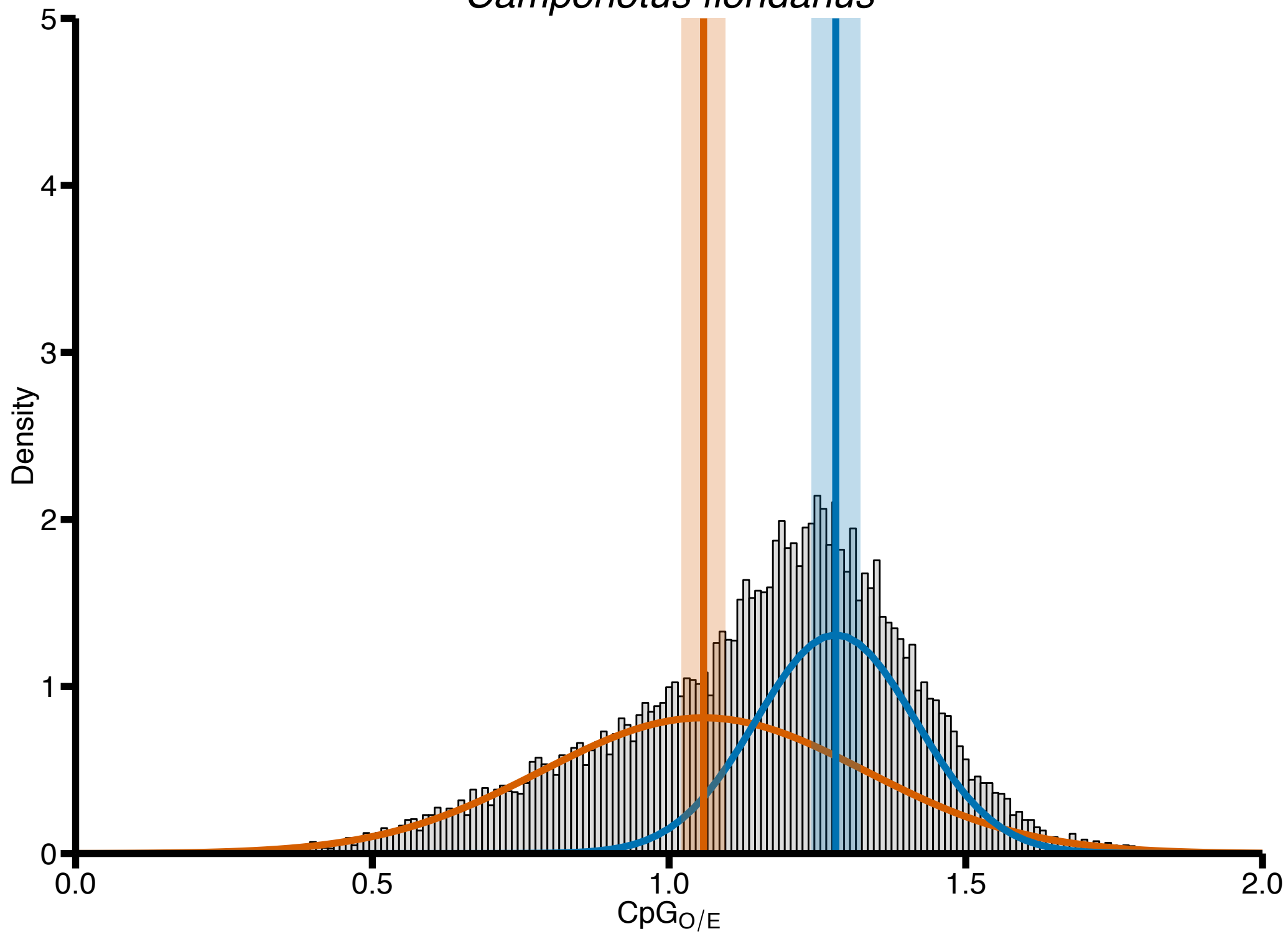
Cerapachys biroi
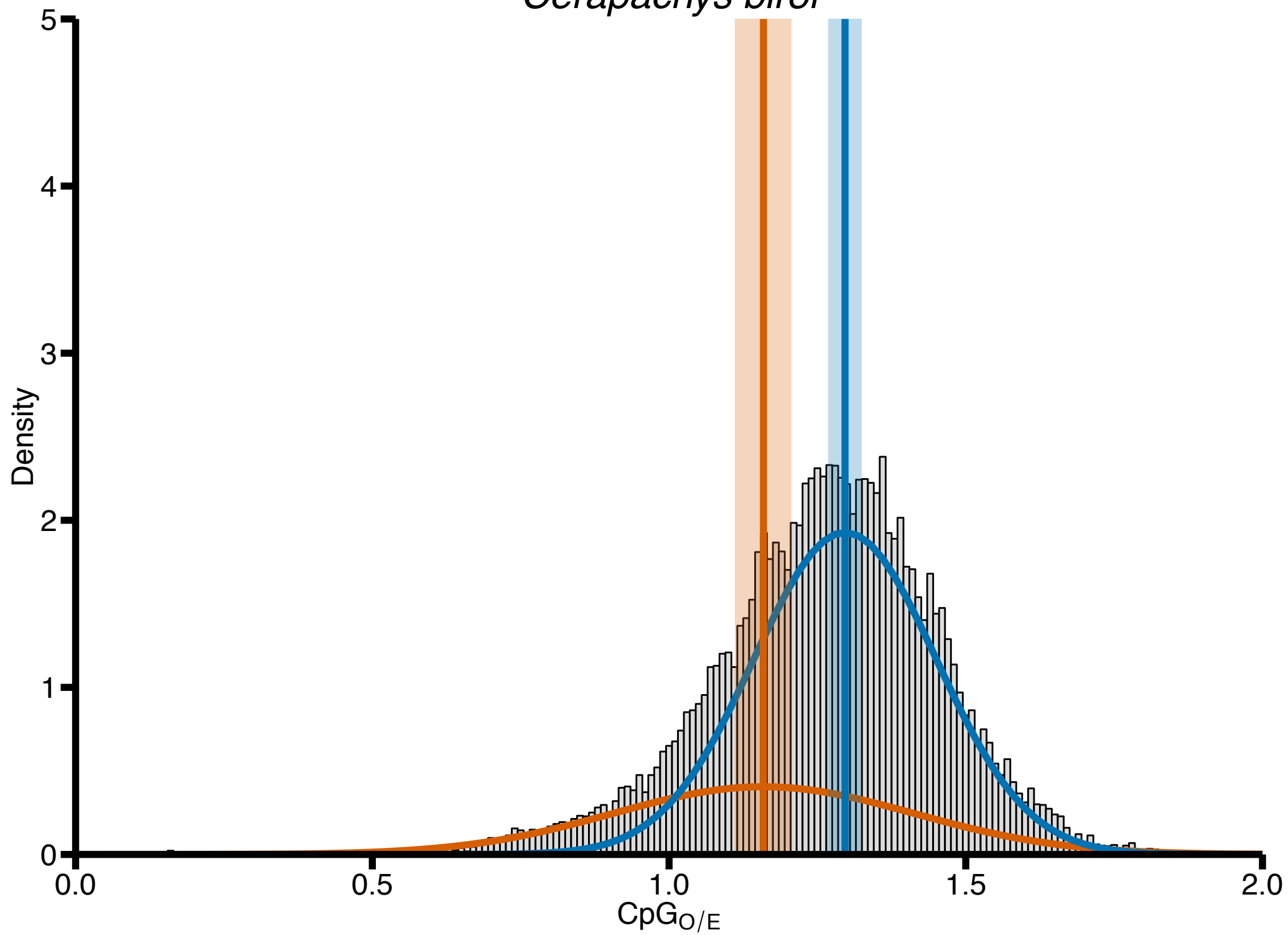
Ceratosolen solmsi marchali
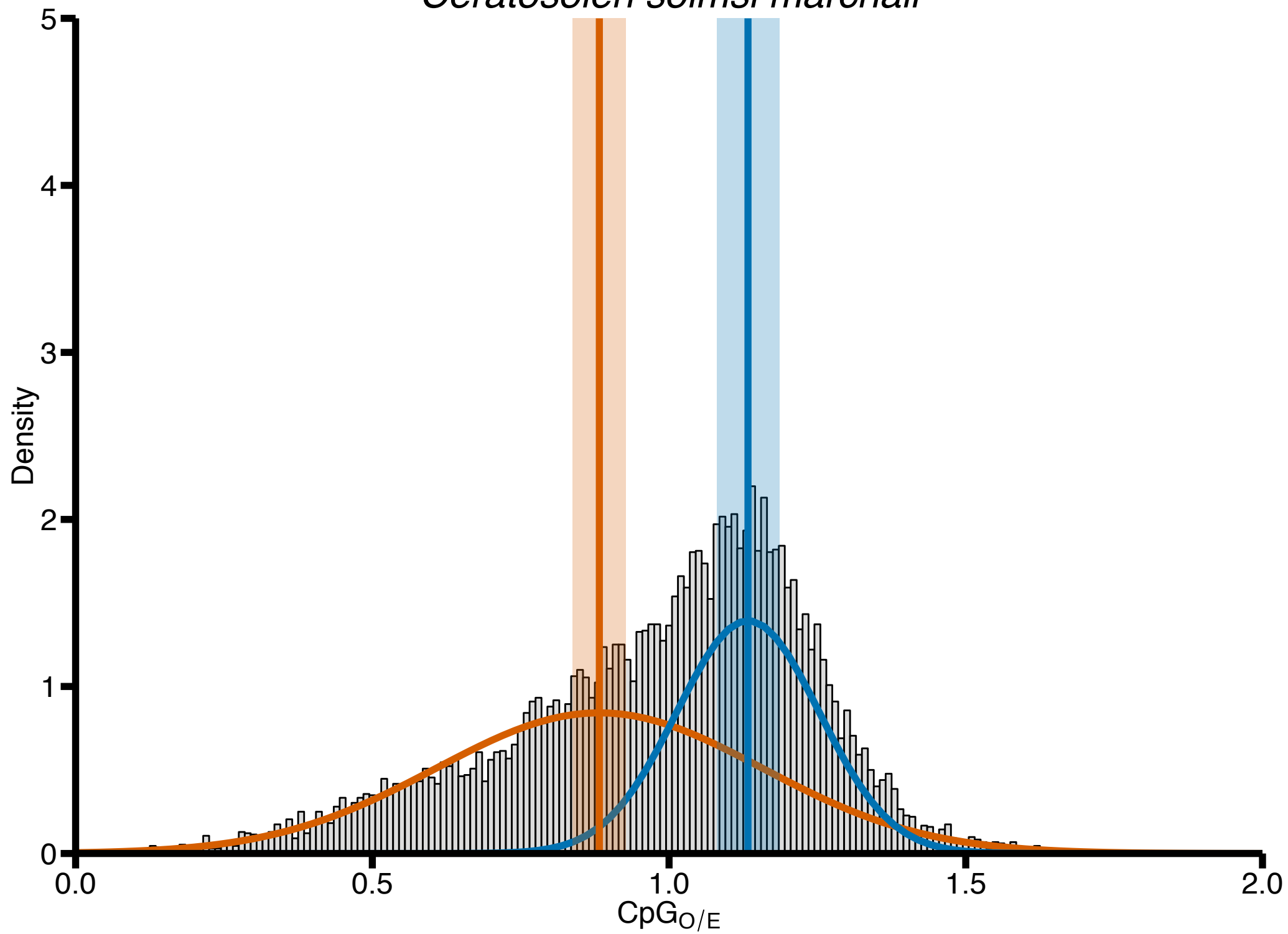
Dinoponera quadriceps
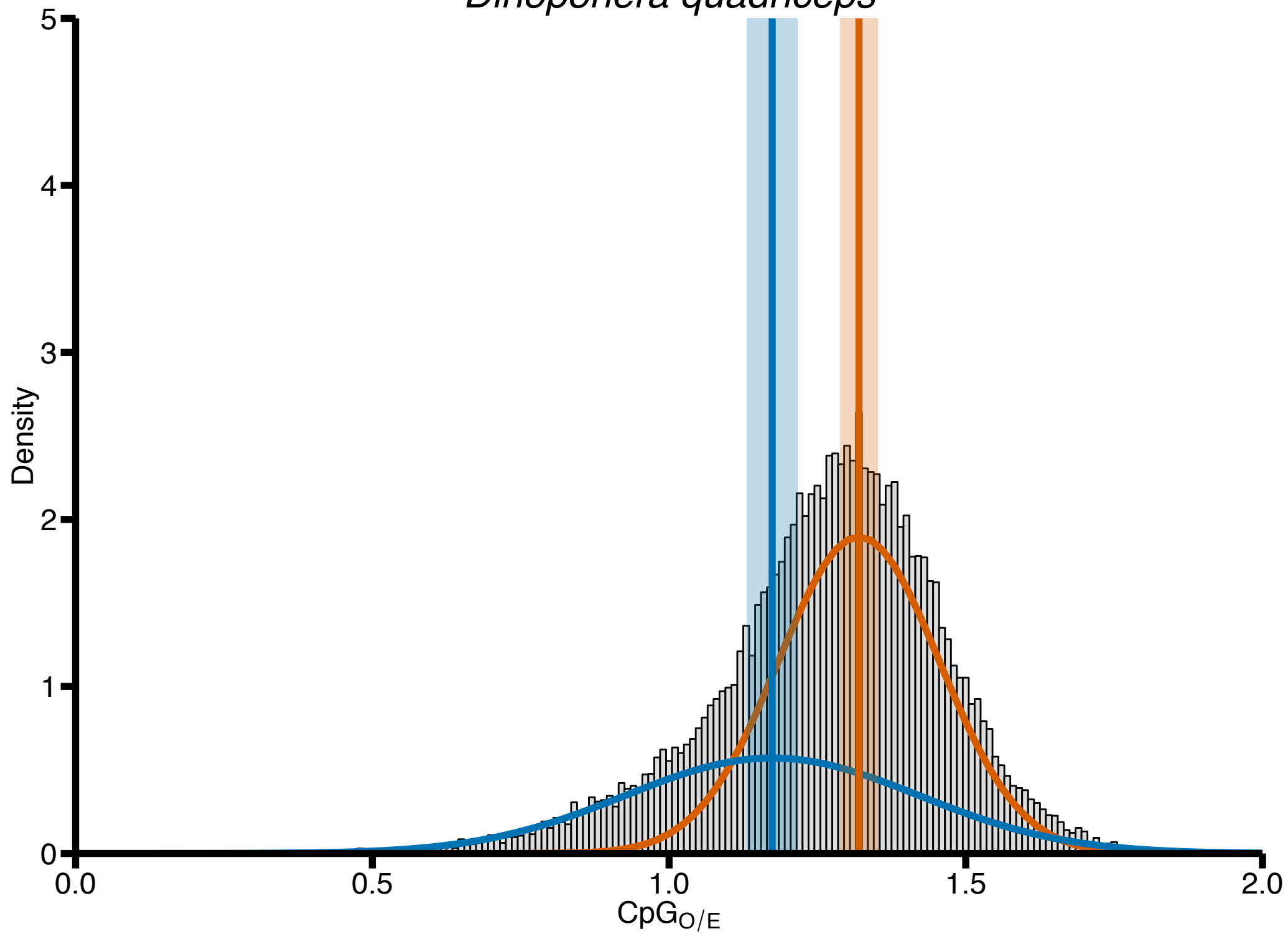
Dufourea novaeangliae
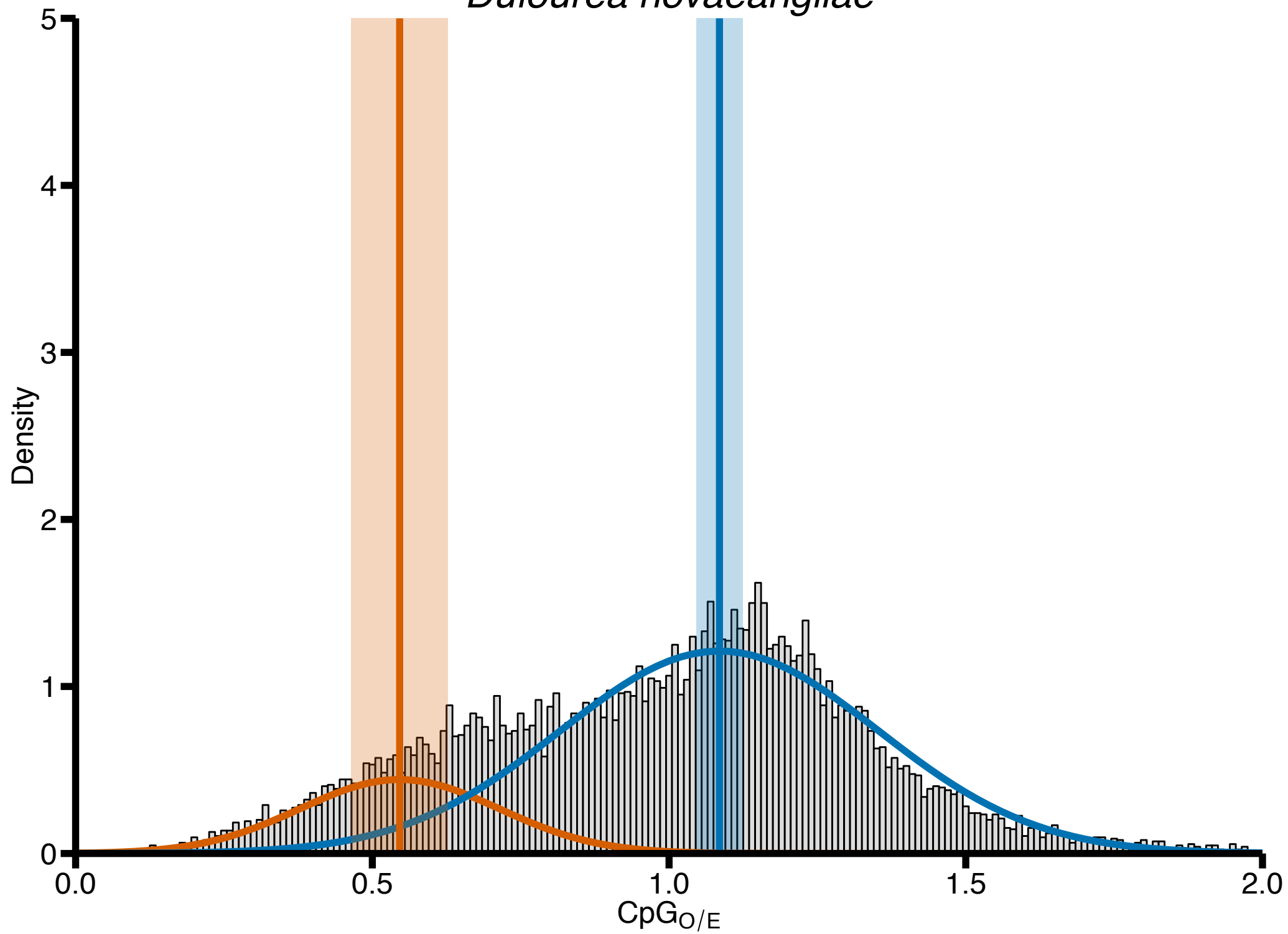
Eufriesea mexicana
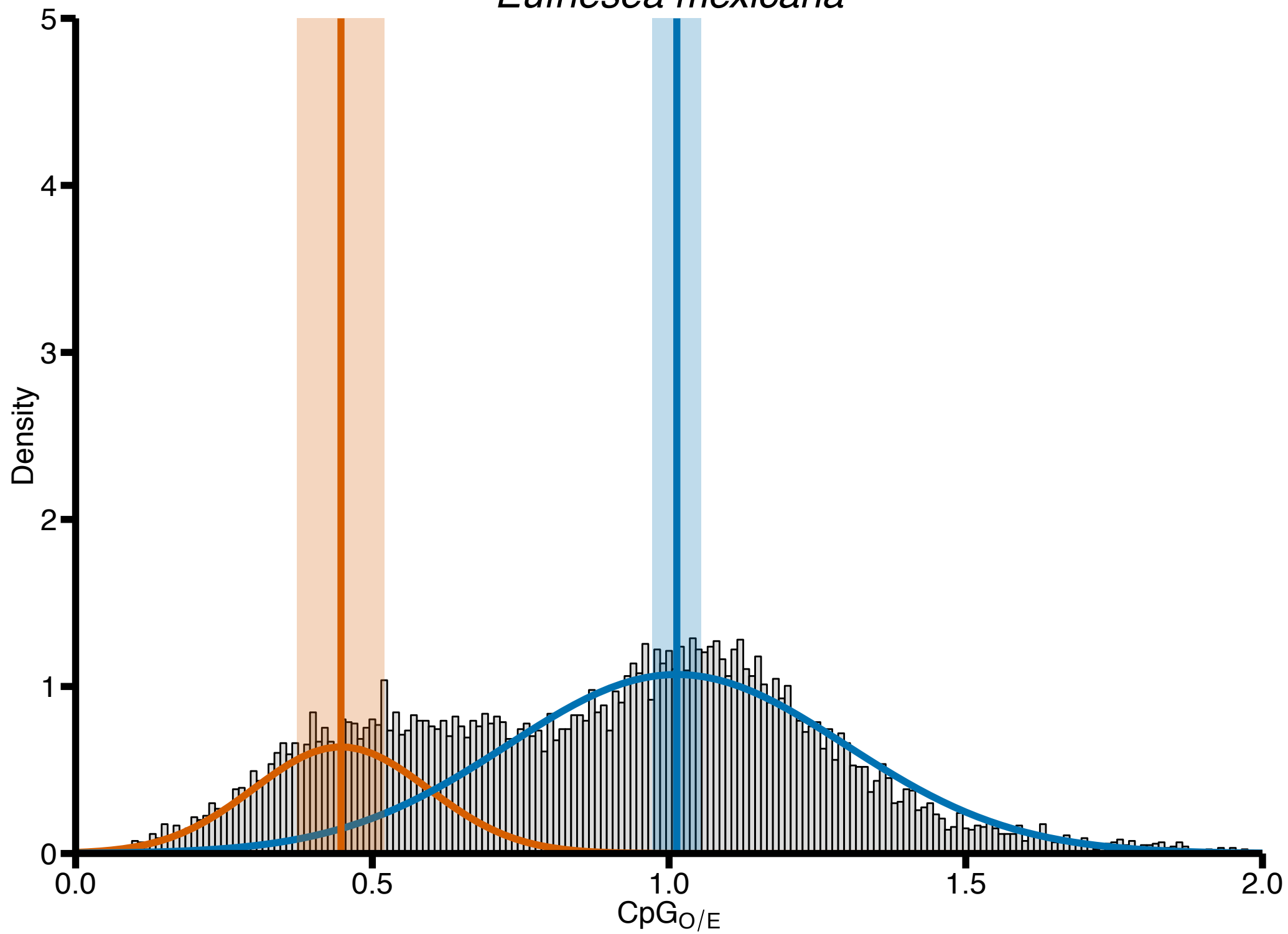
Fopius arisanus
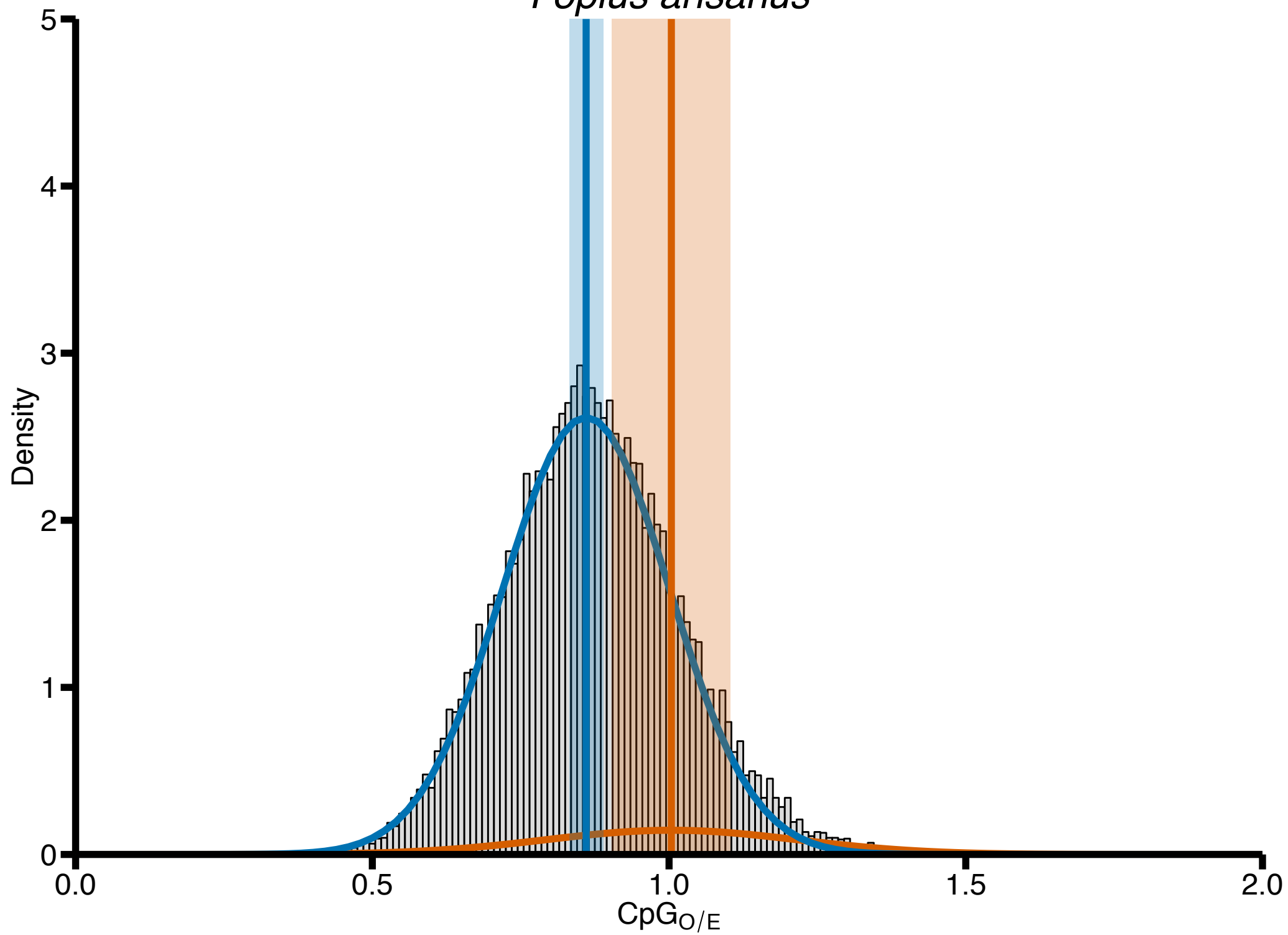
Habropoda laboriosa
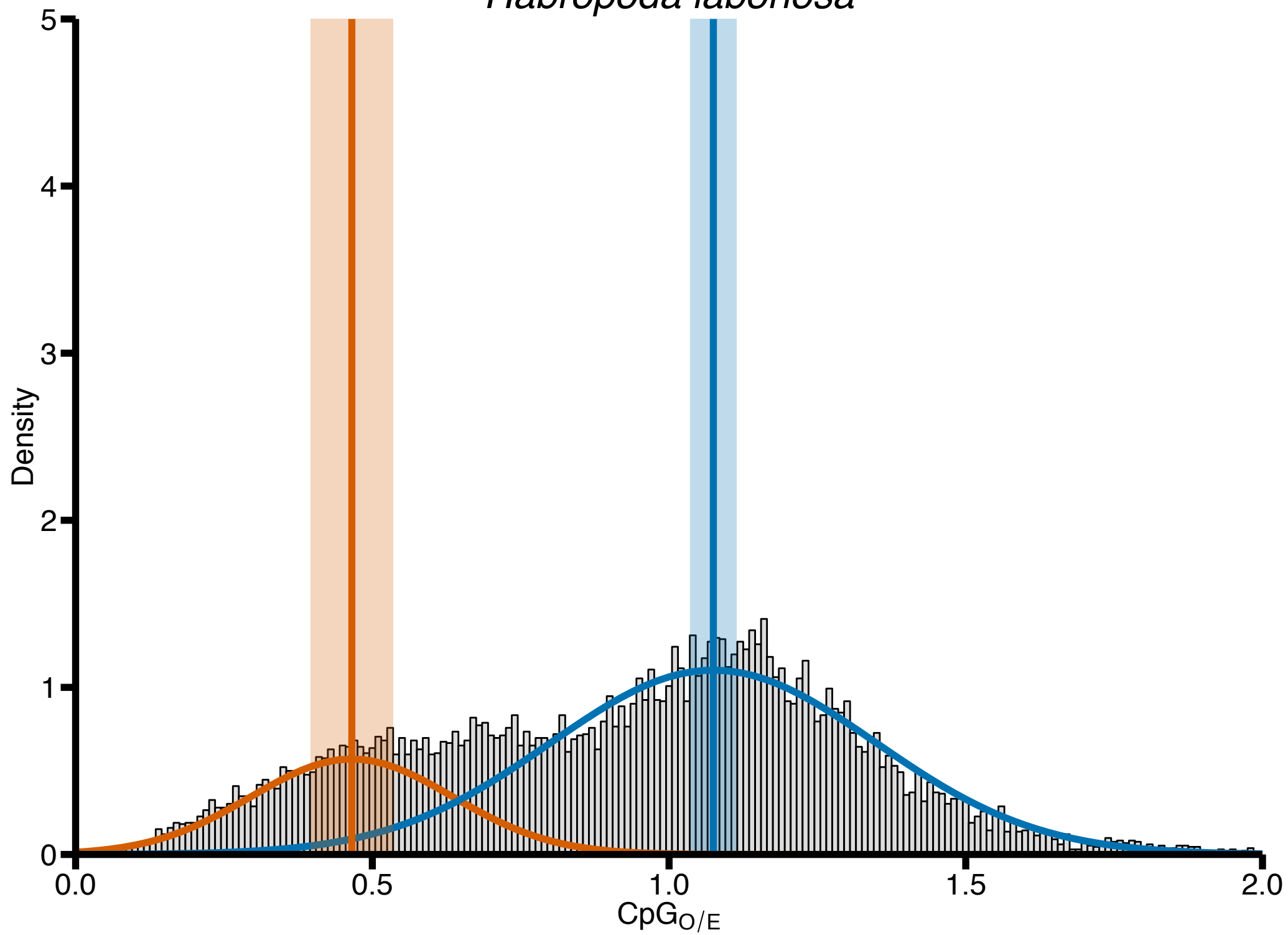
Harpegnathos saltator
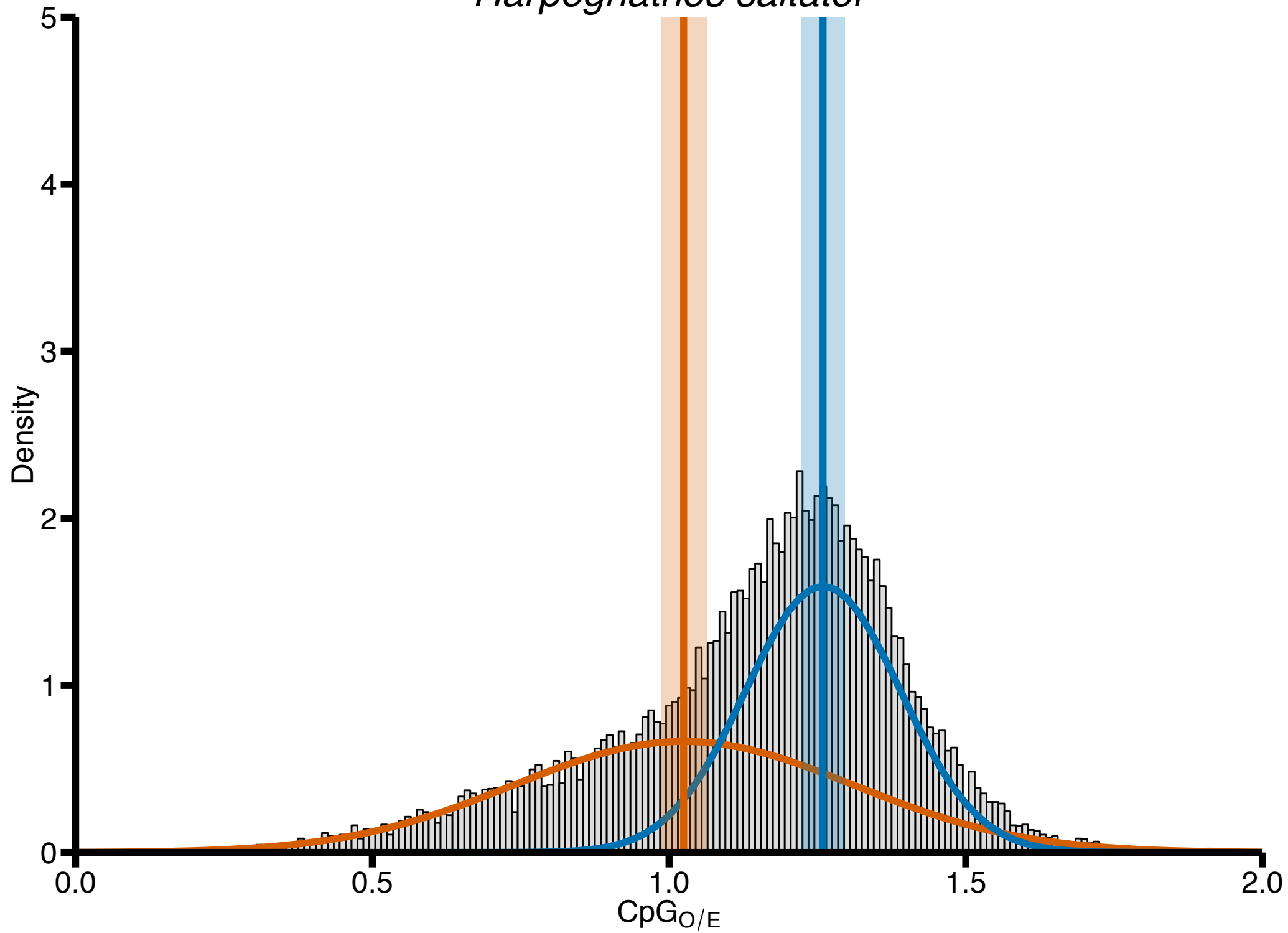
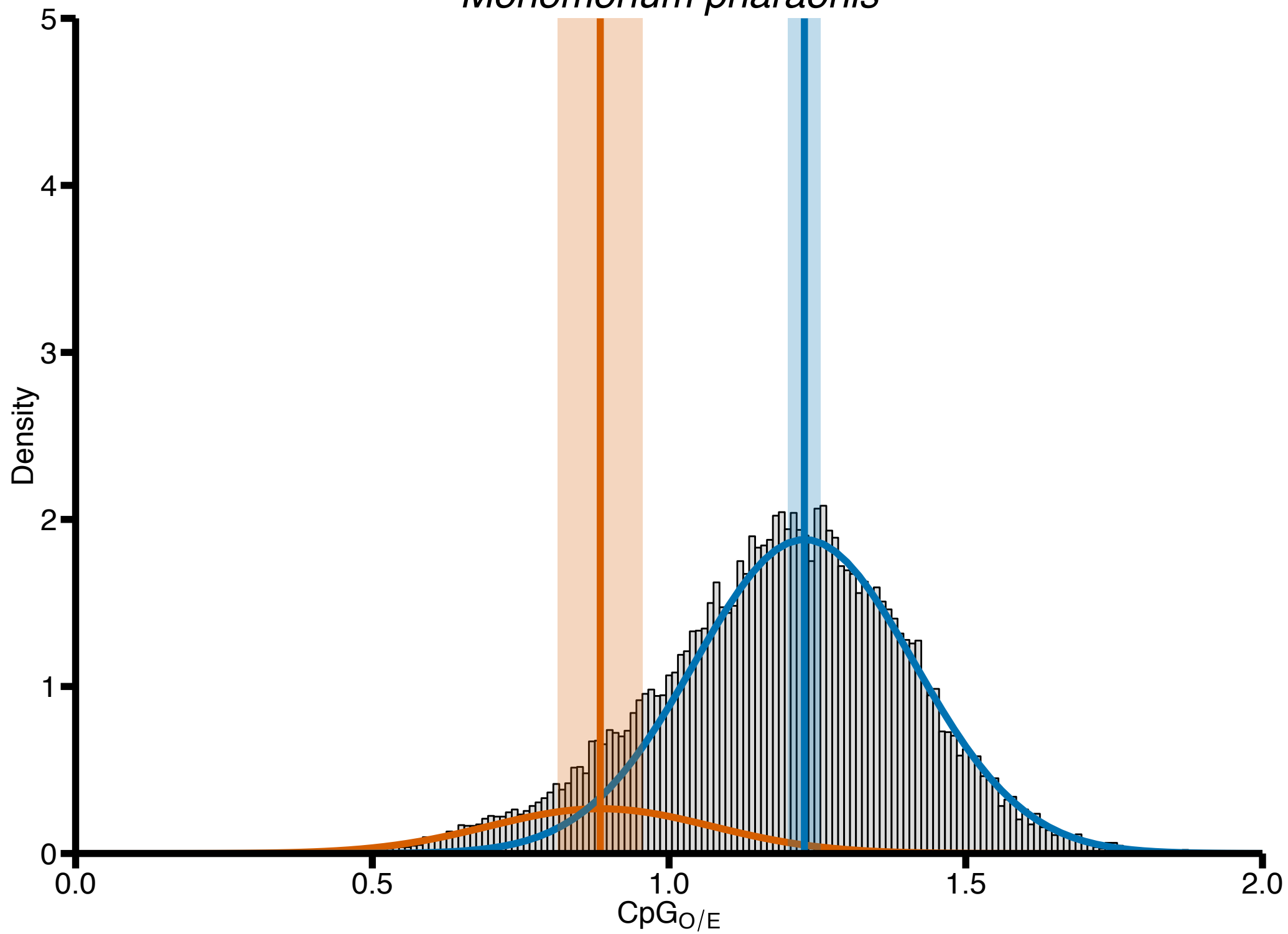*Lasioglossum albipes*
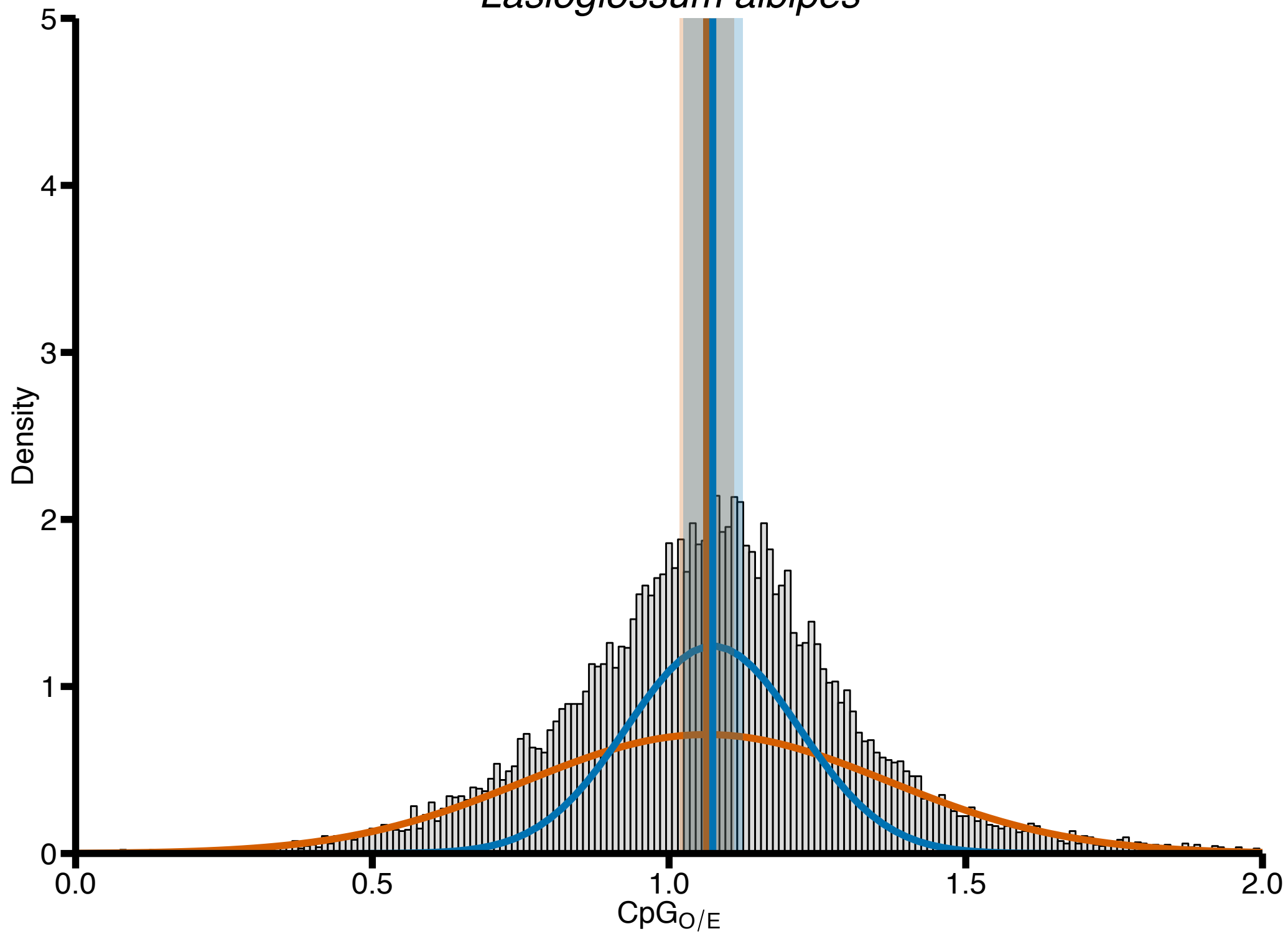
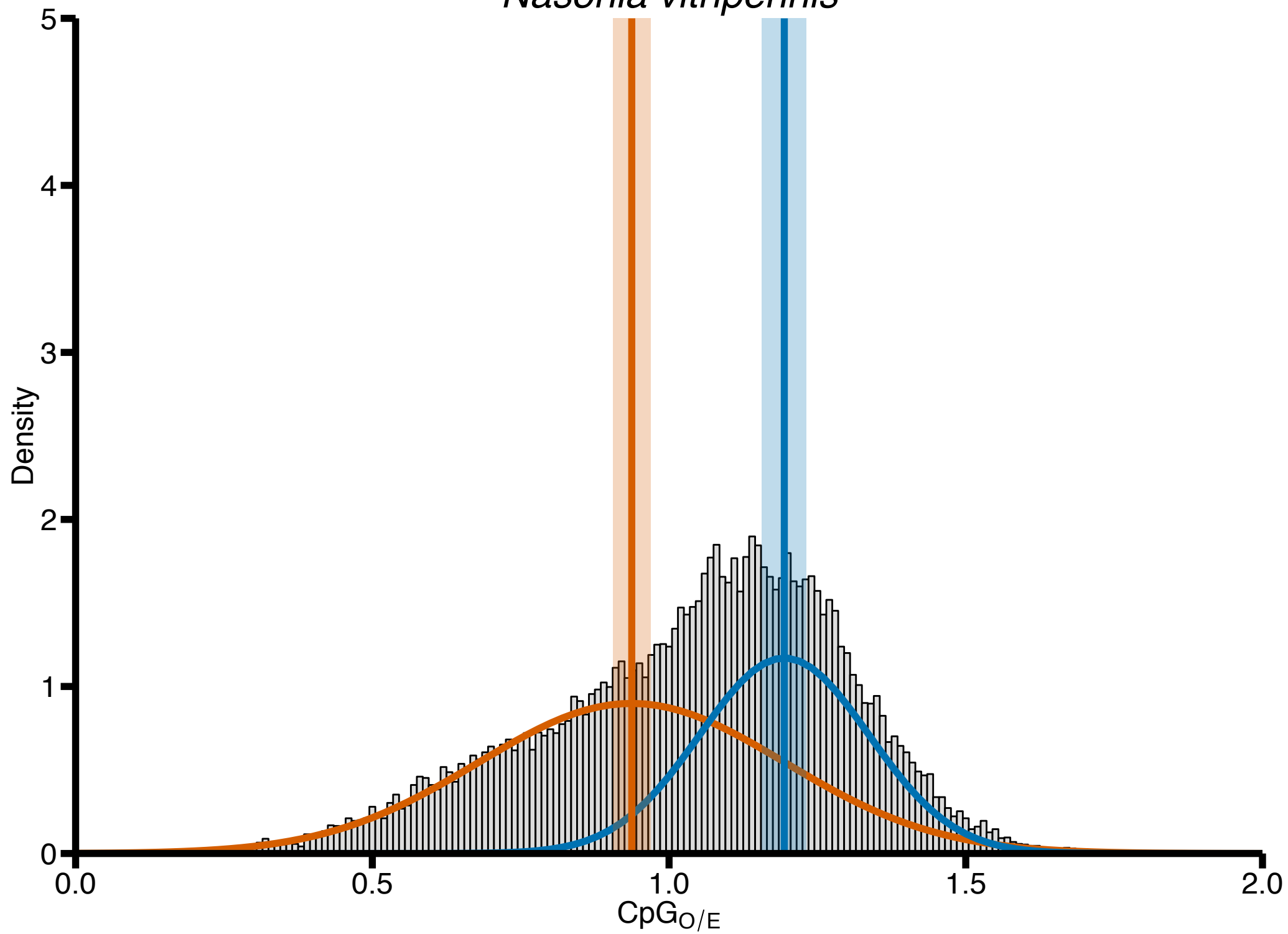*Linepithema humile*
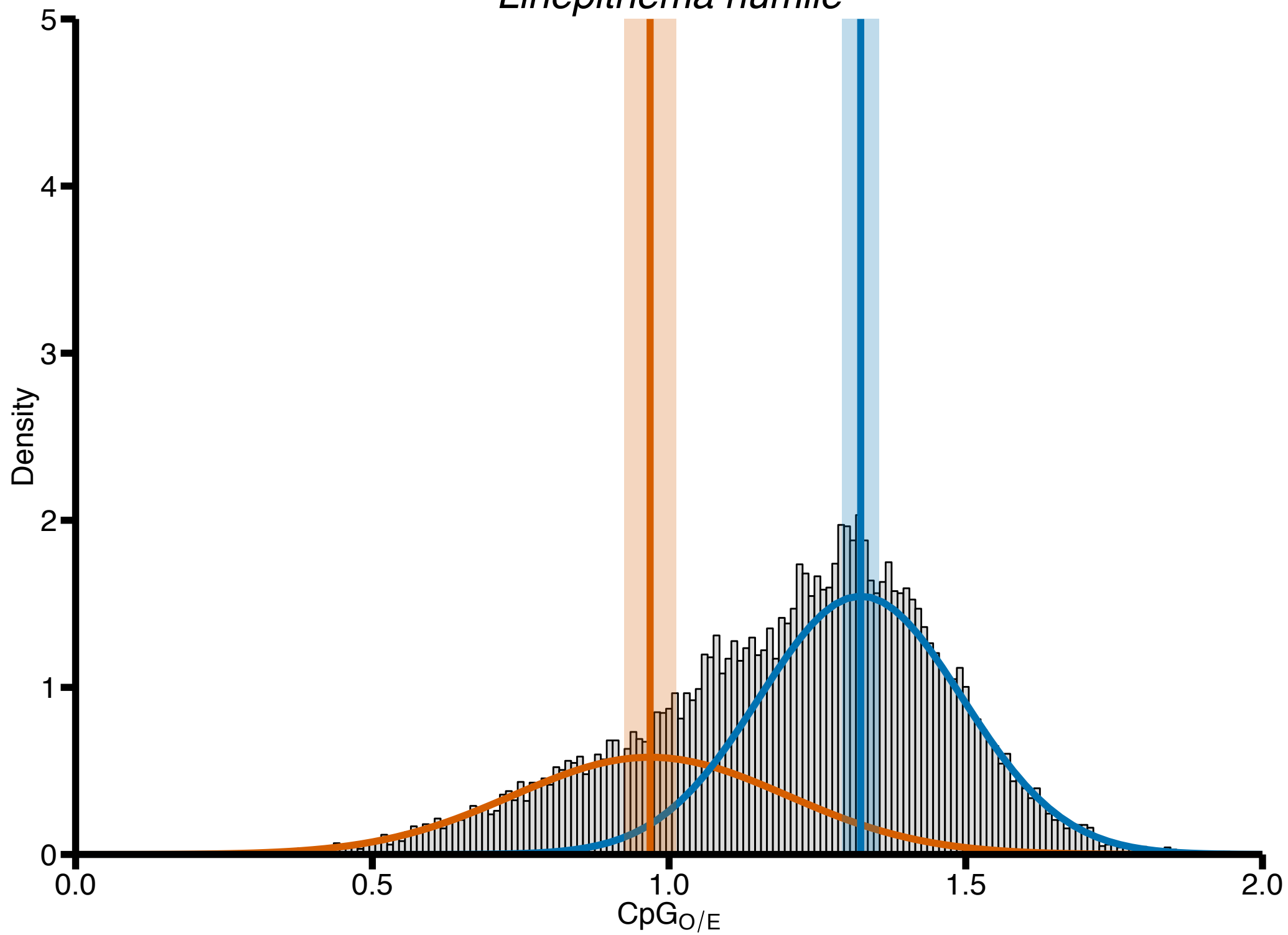
Megachile rotundata
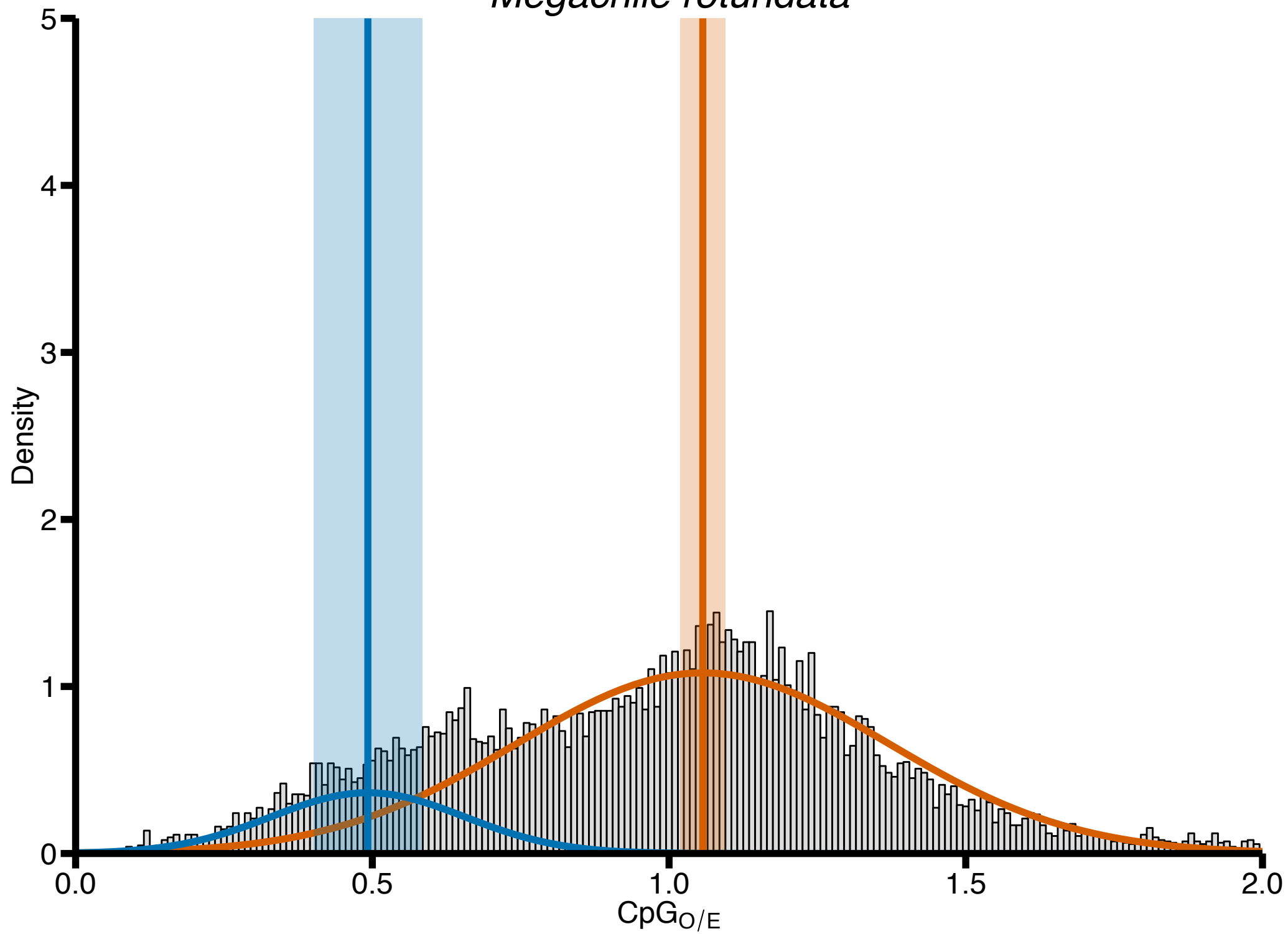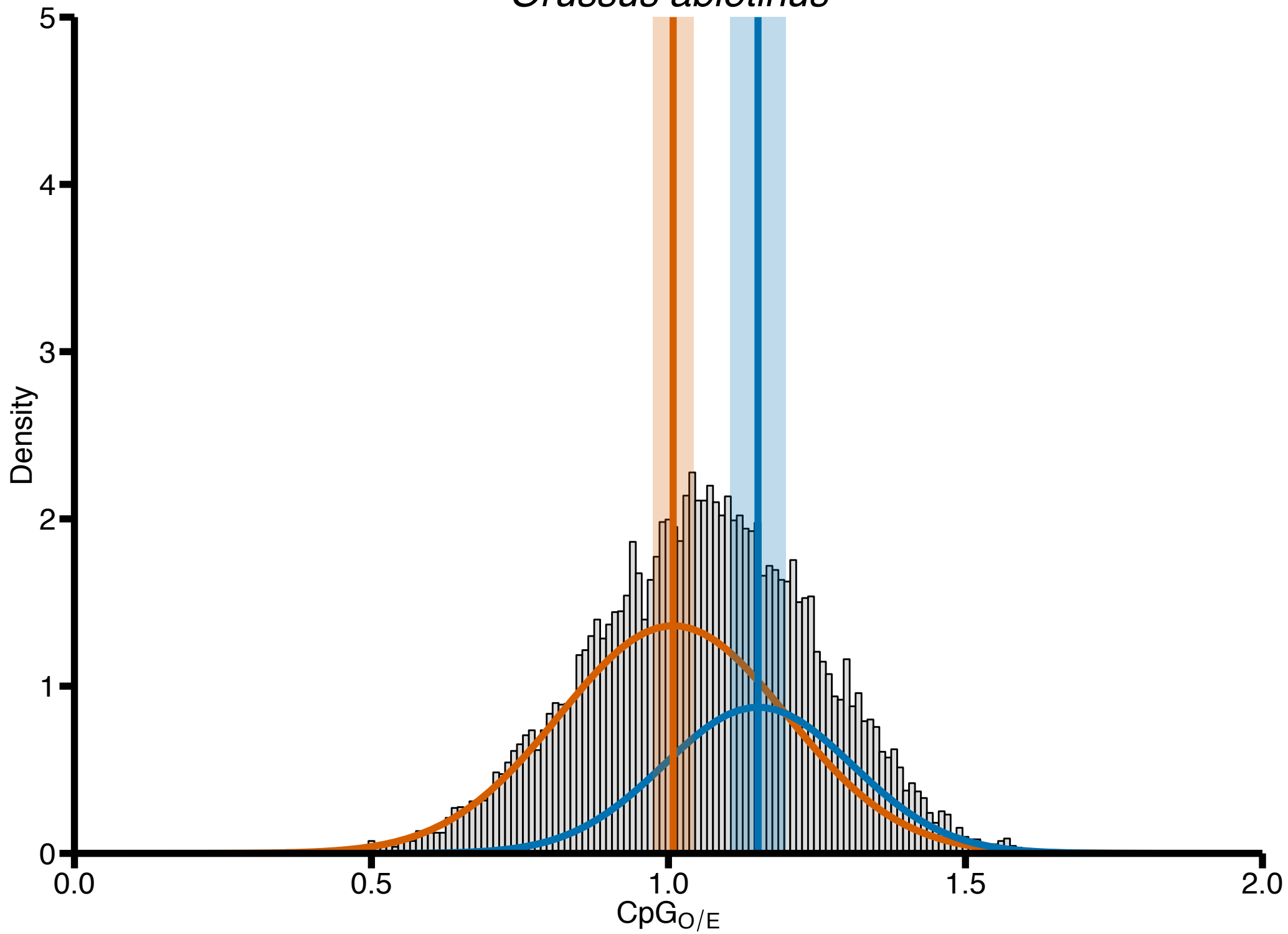
Melipona quadrifasciata
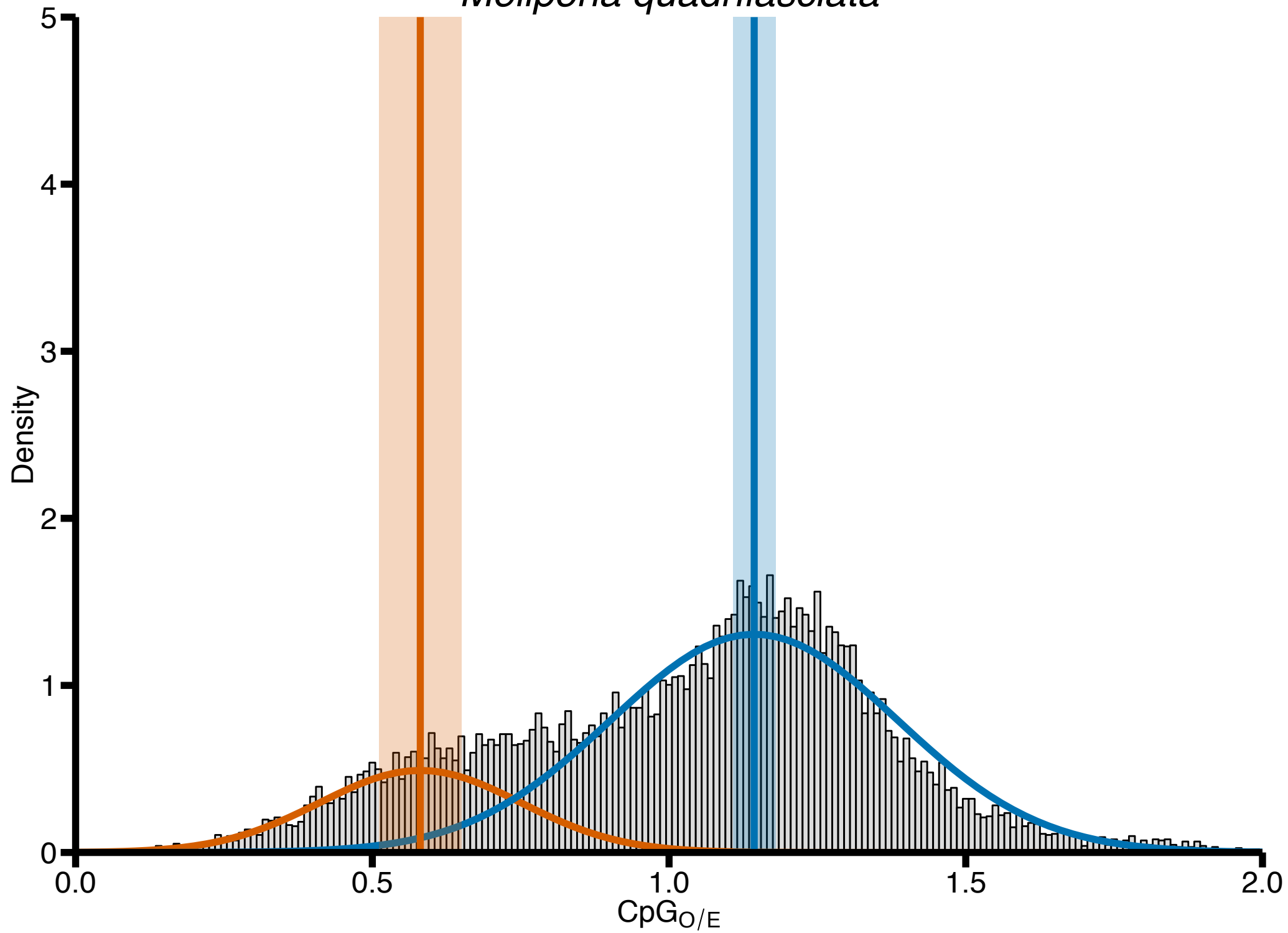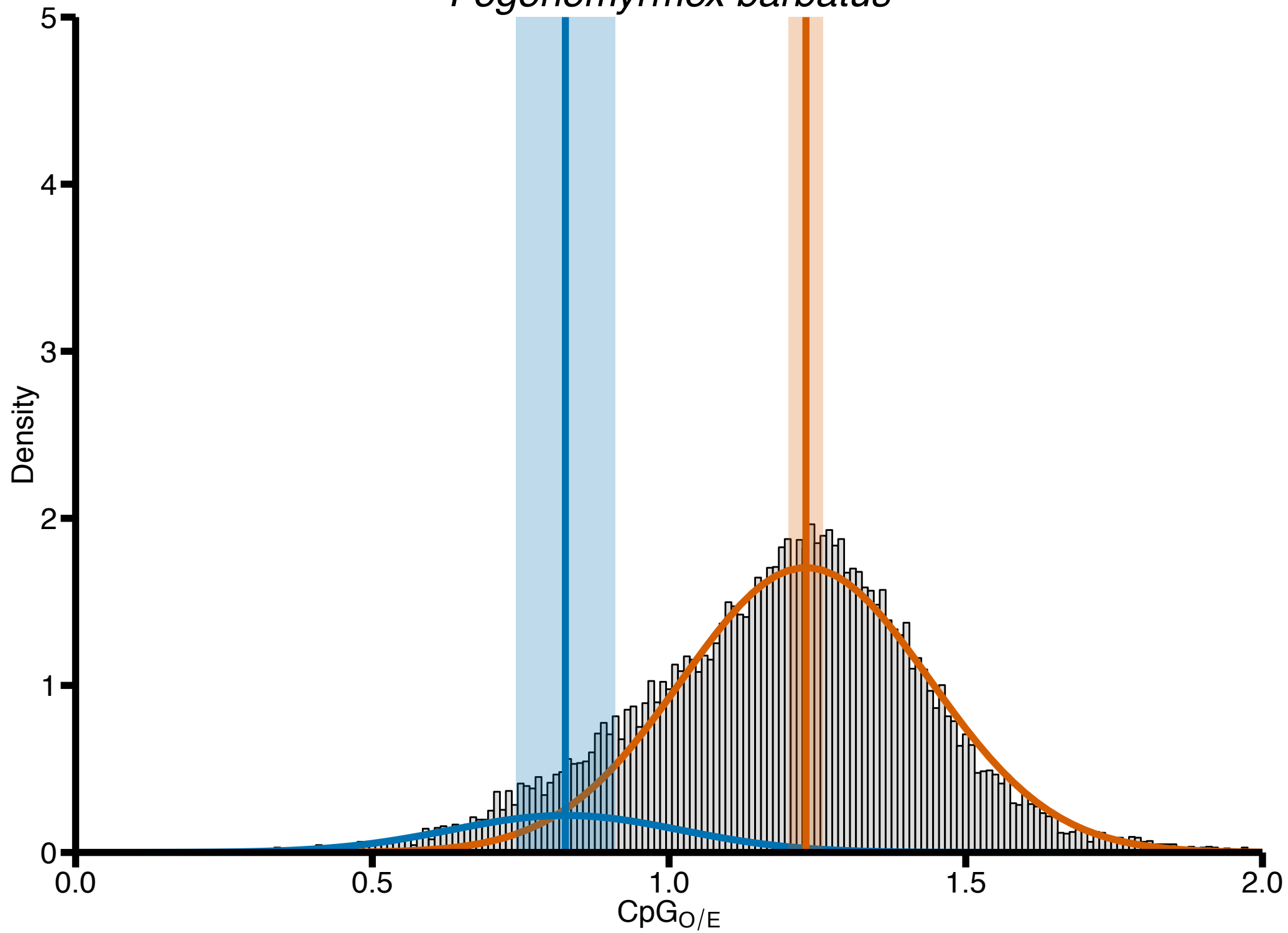
Microplitis demolitor
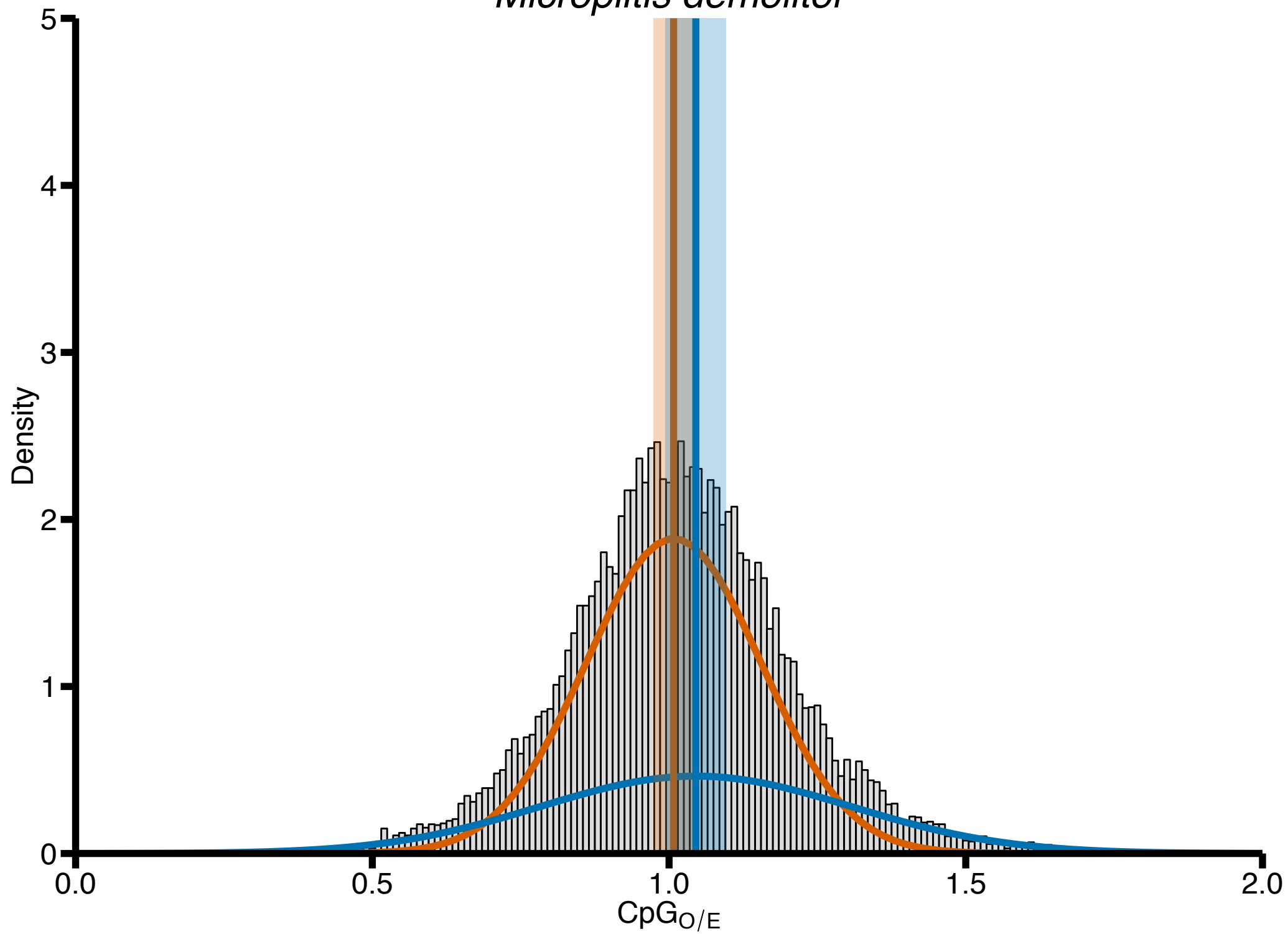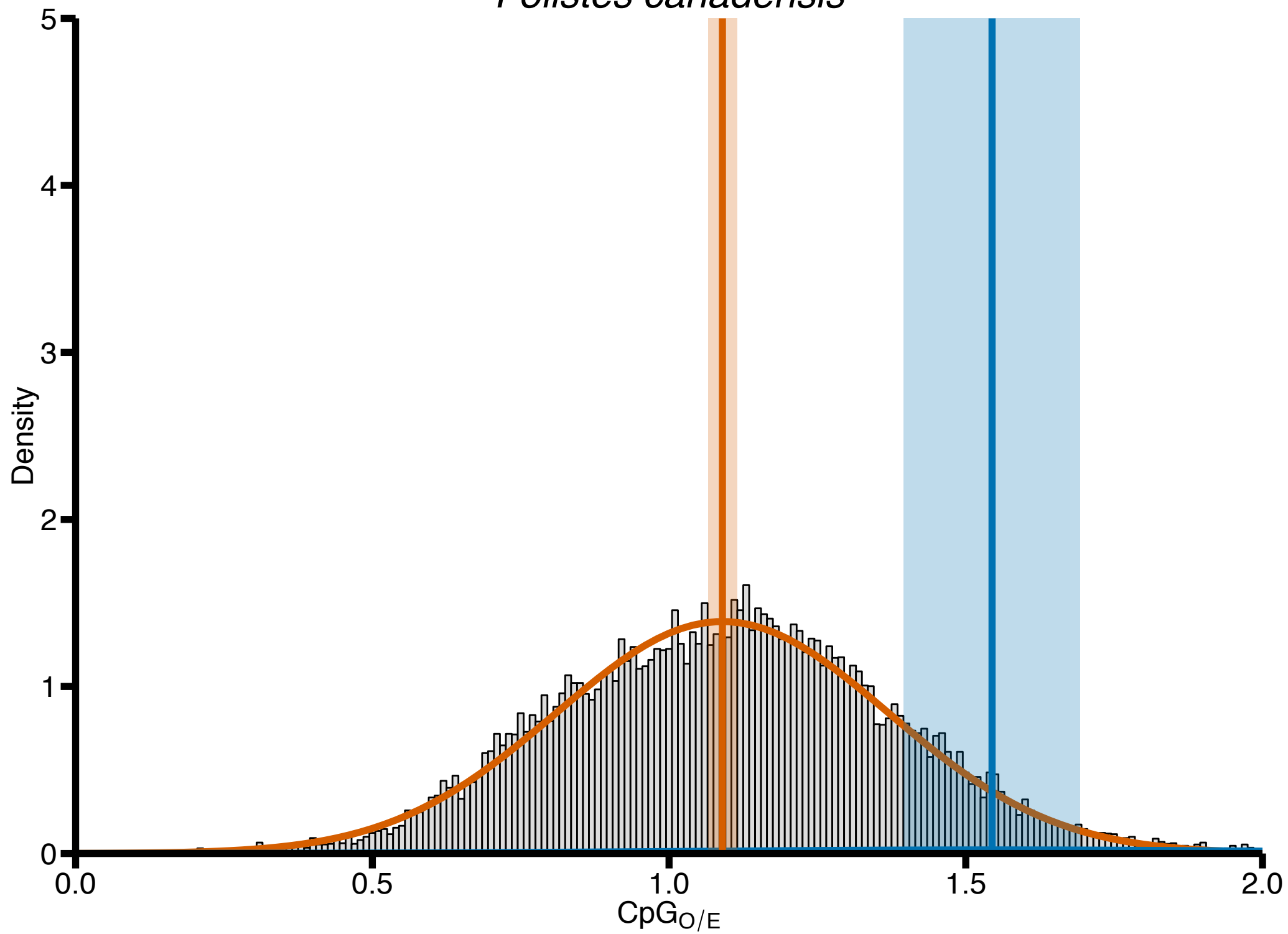
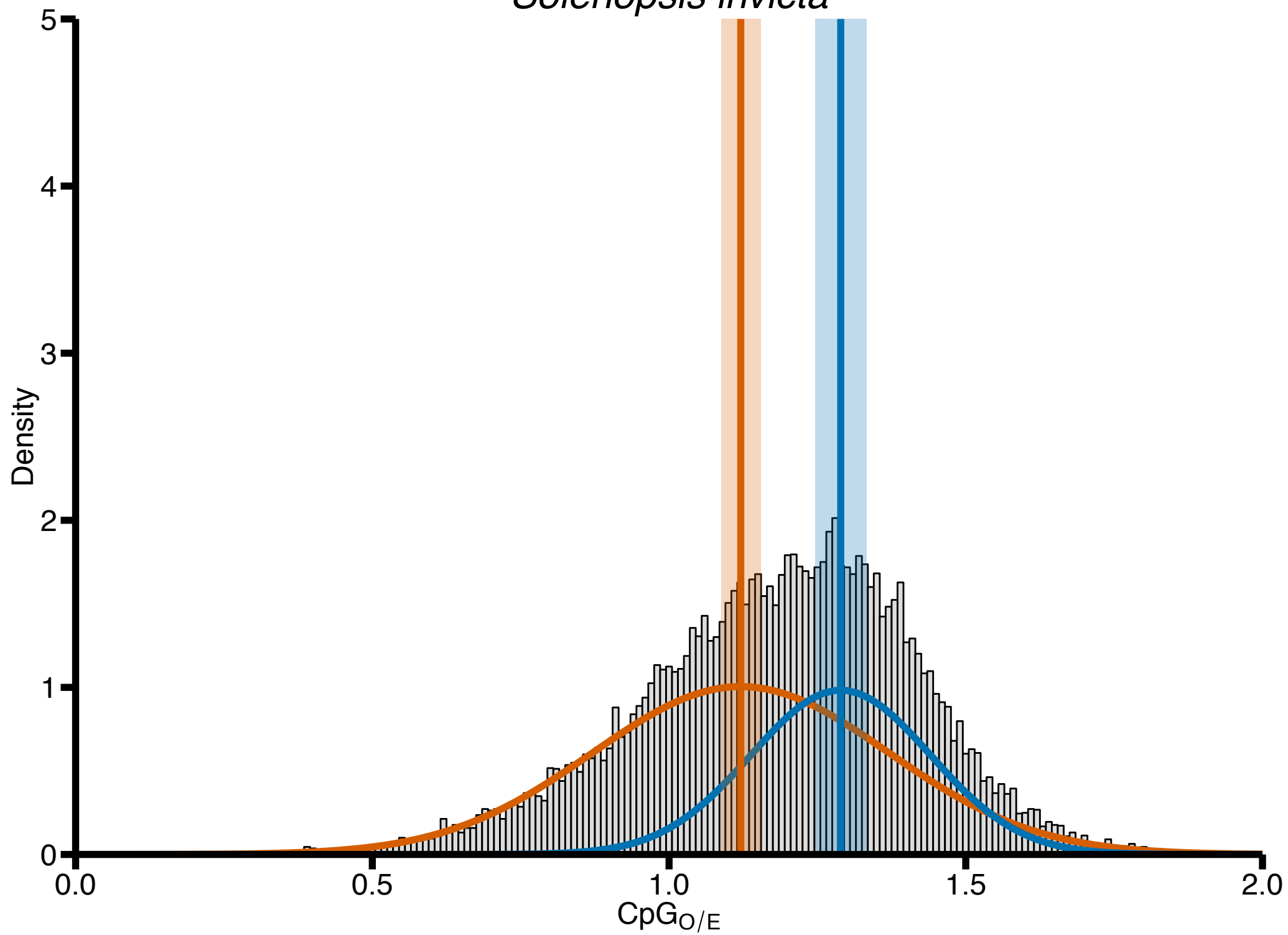*Monomorium pharaonis*
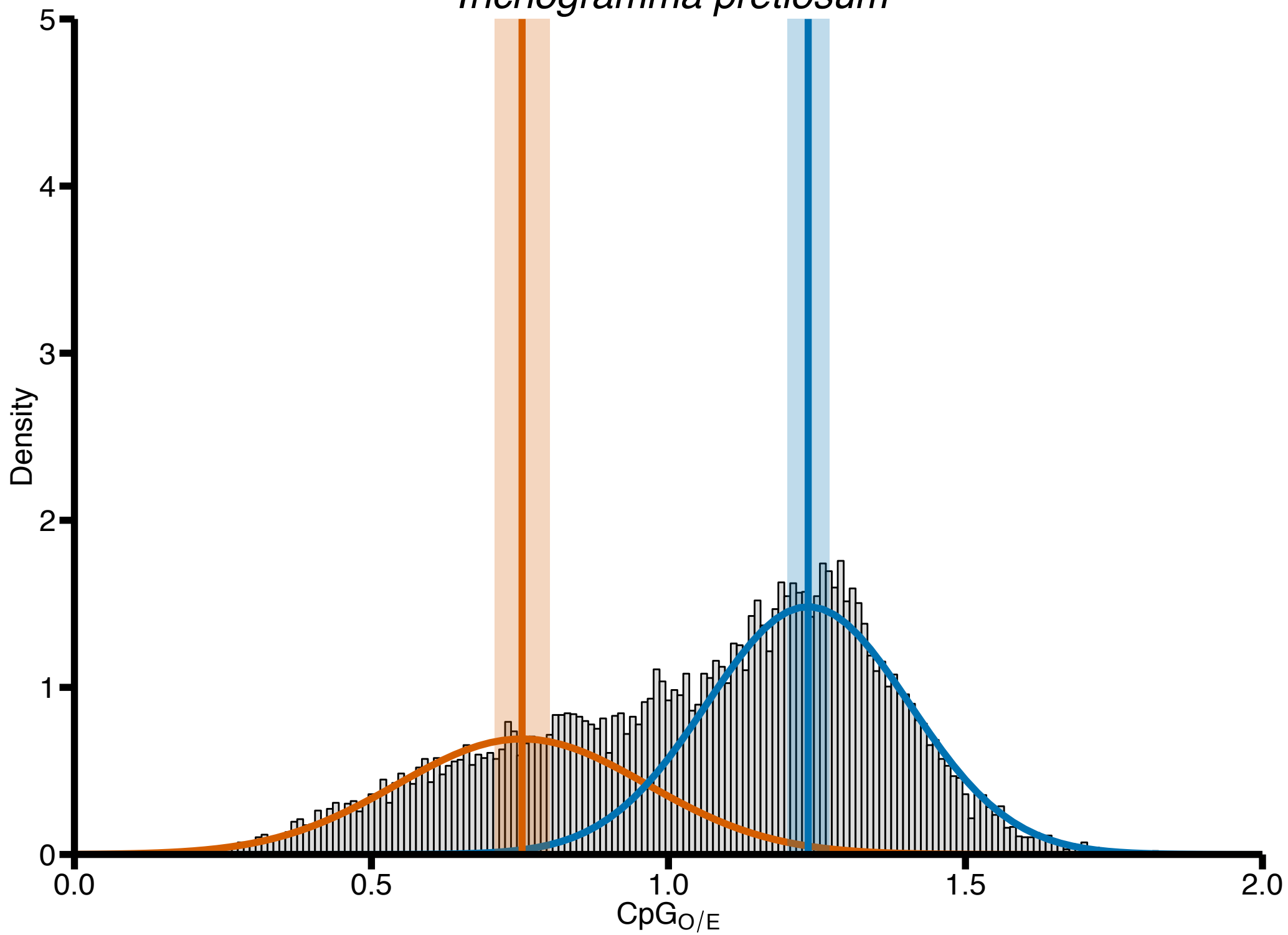
Nasonia vitripennis

Orussus abietinus
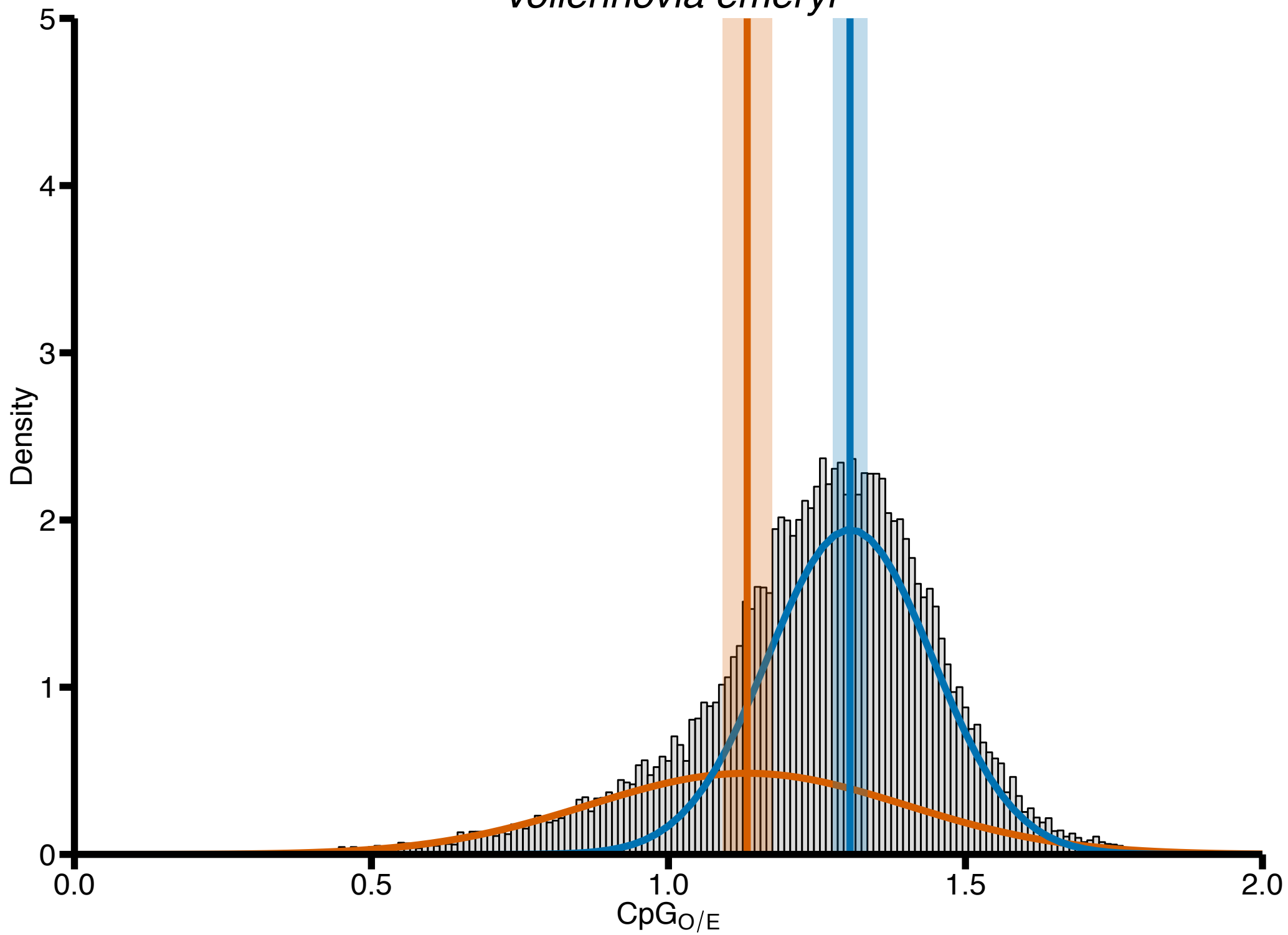
Pogonomyrmex barbatus

Polistes canadensis

Solenopsis invicta

Trichogramma pretiosum

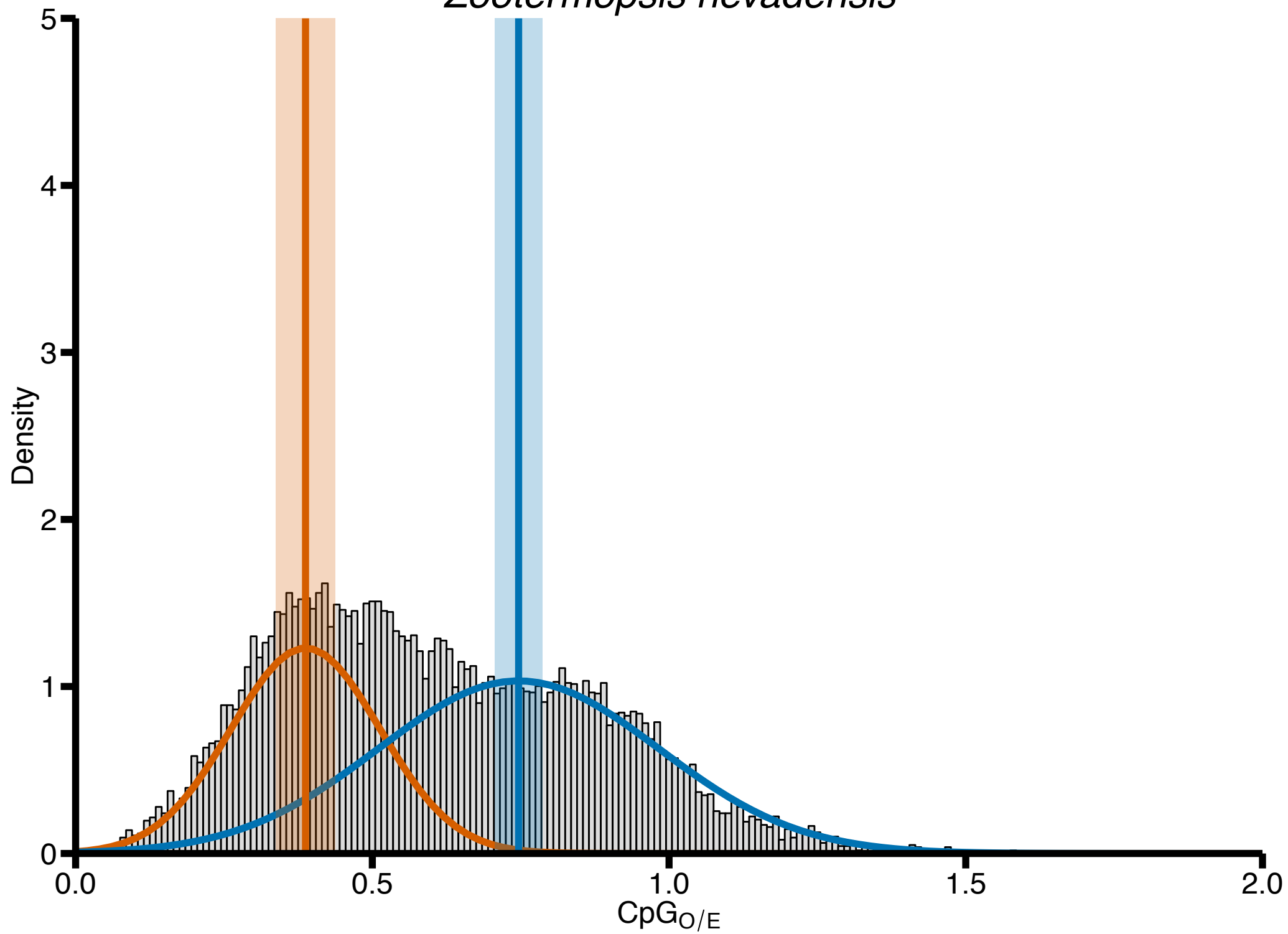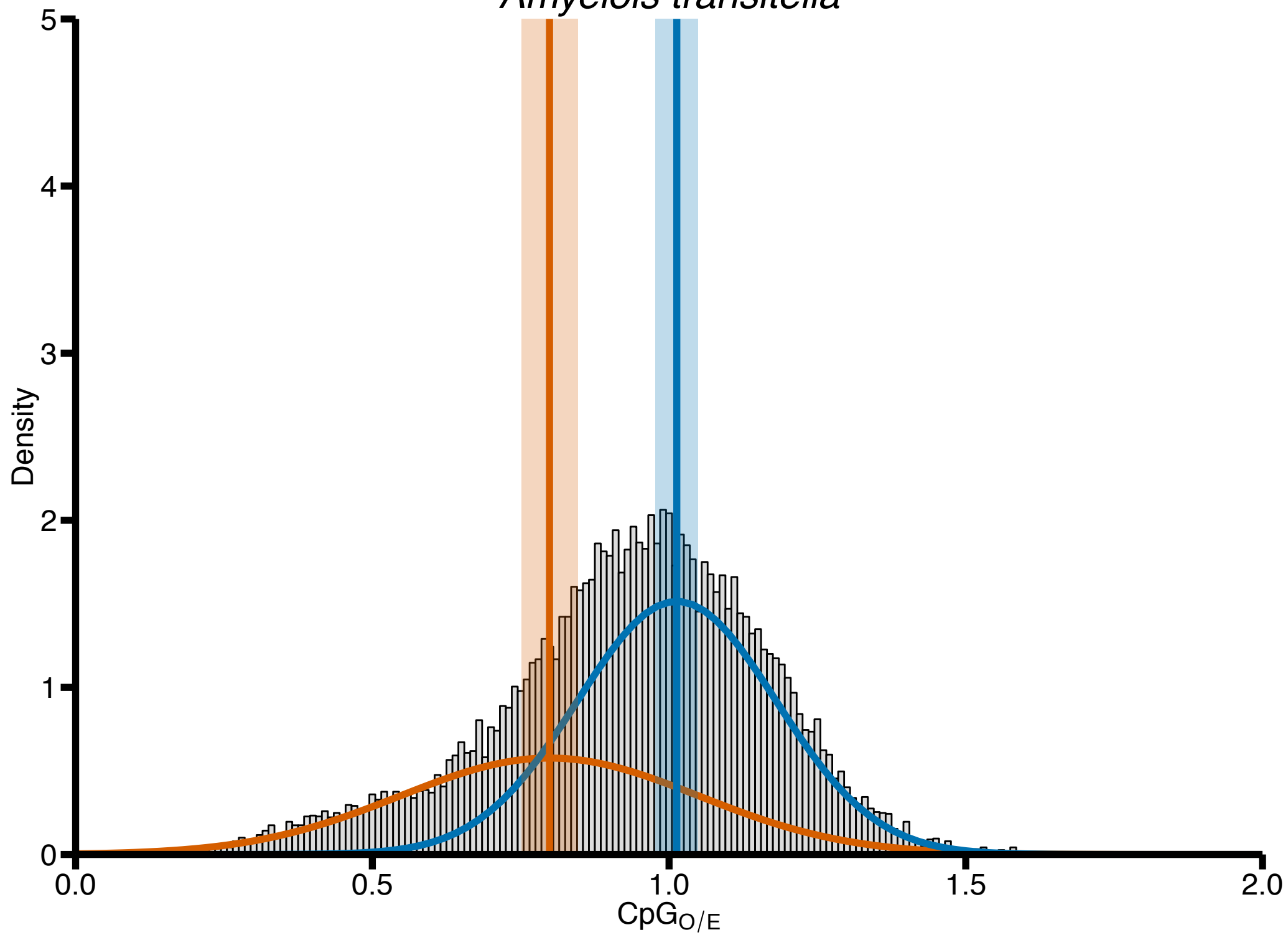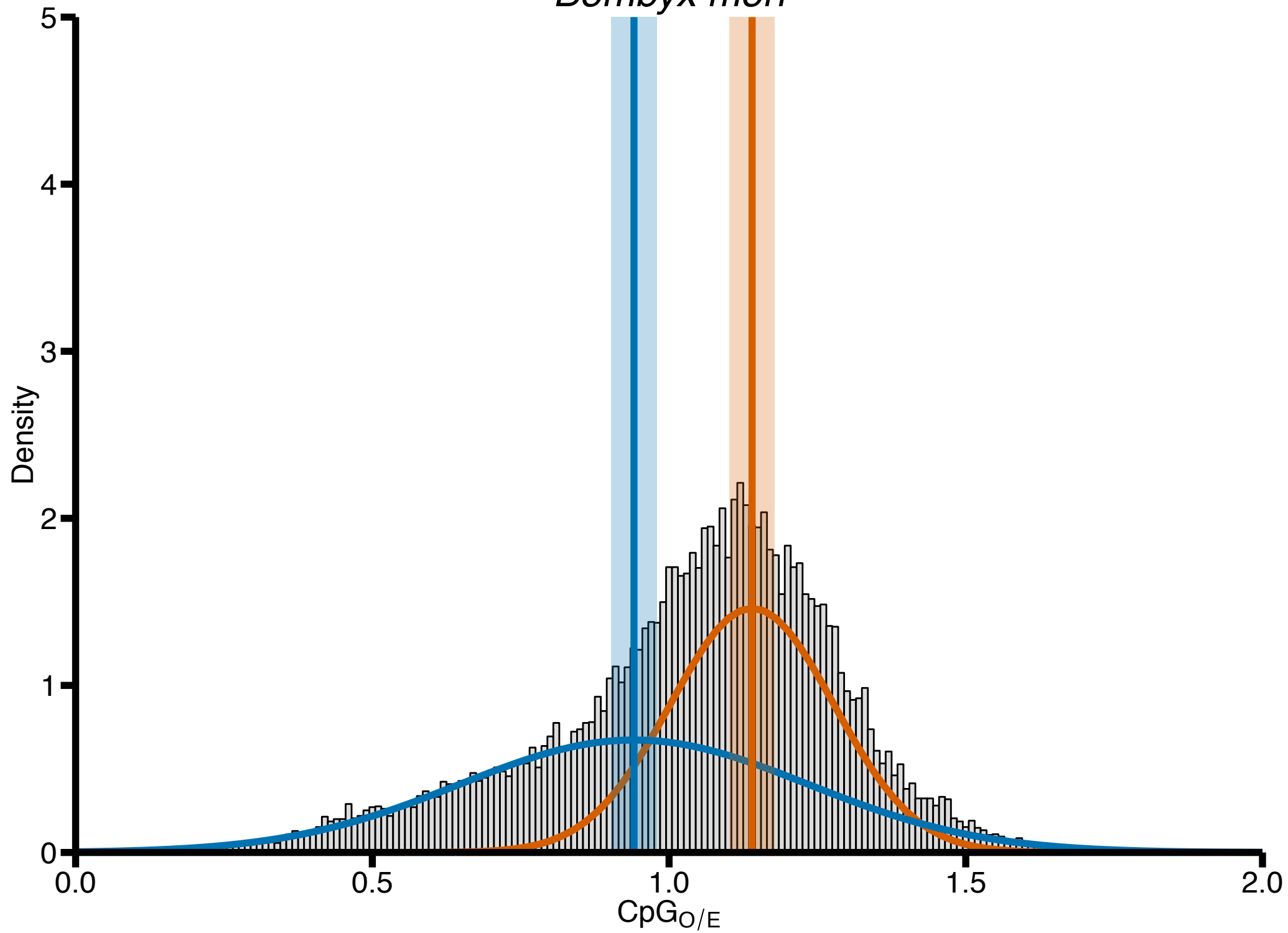*Vollenhovia emeryi*

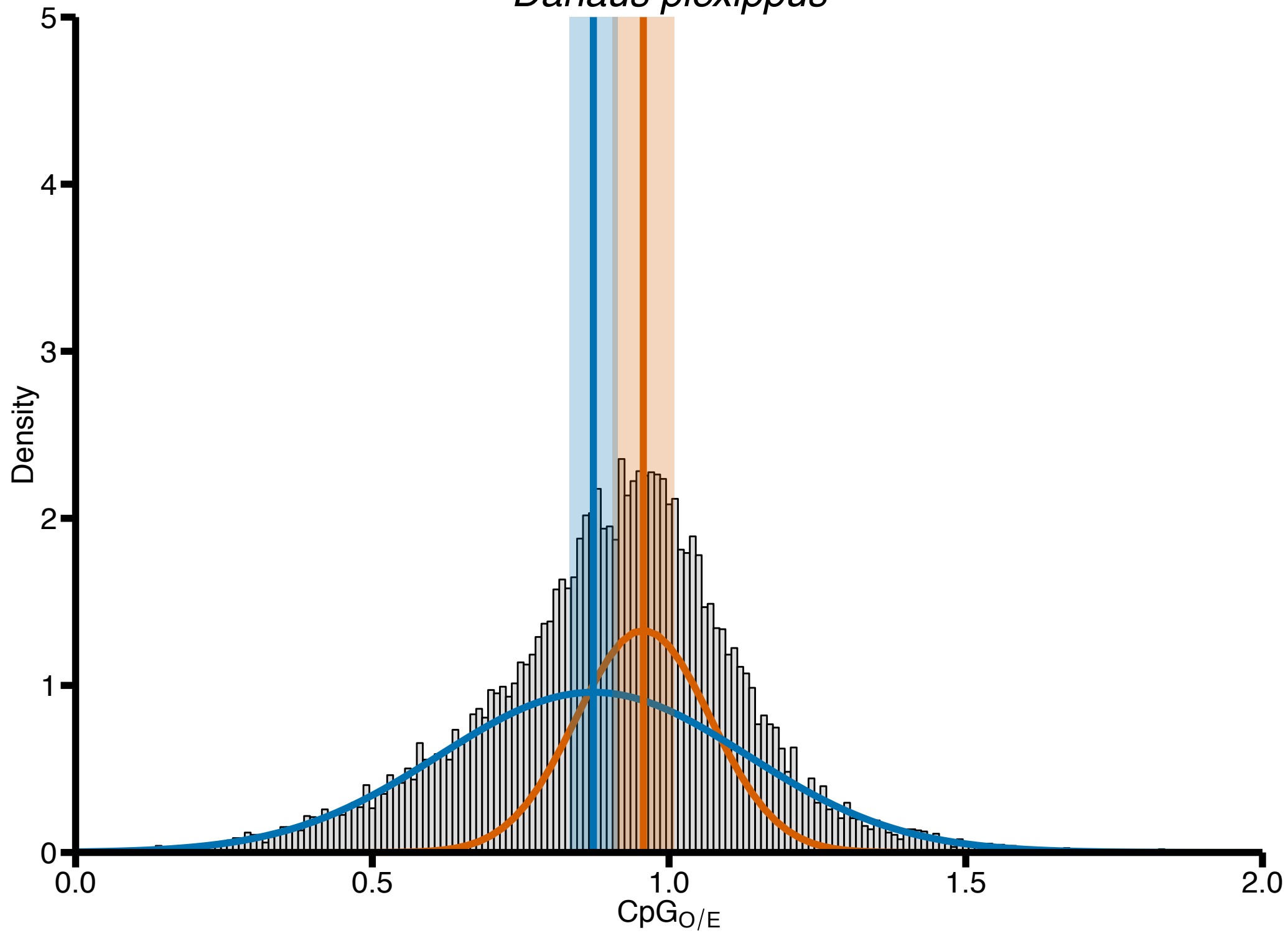*Wasmannia auropunctata*
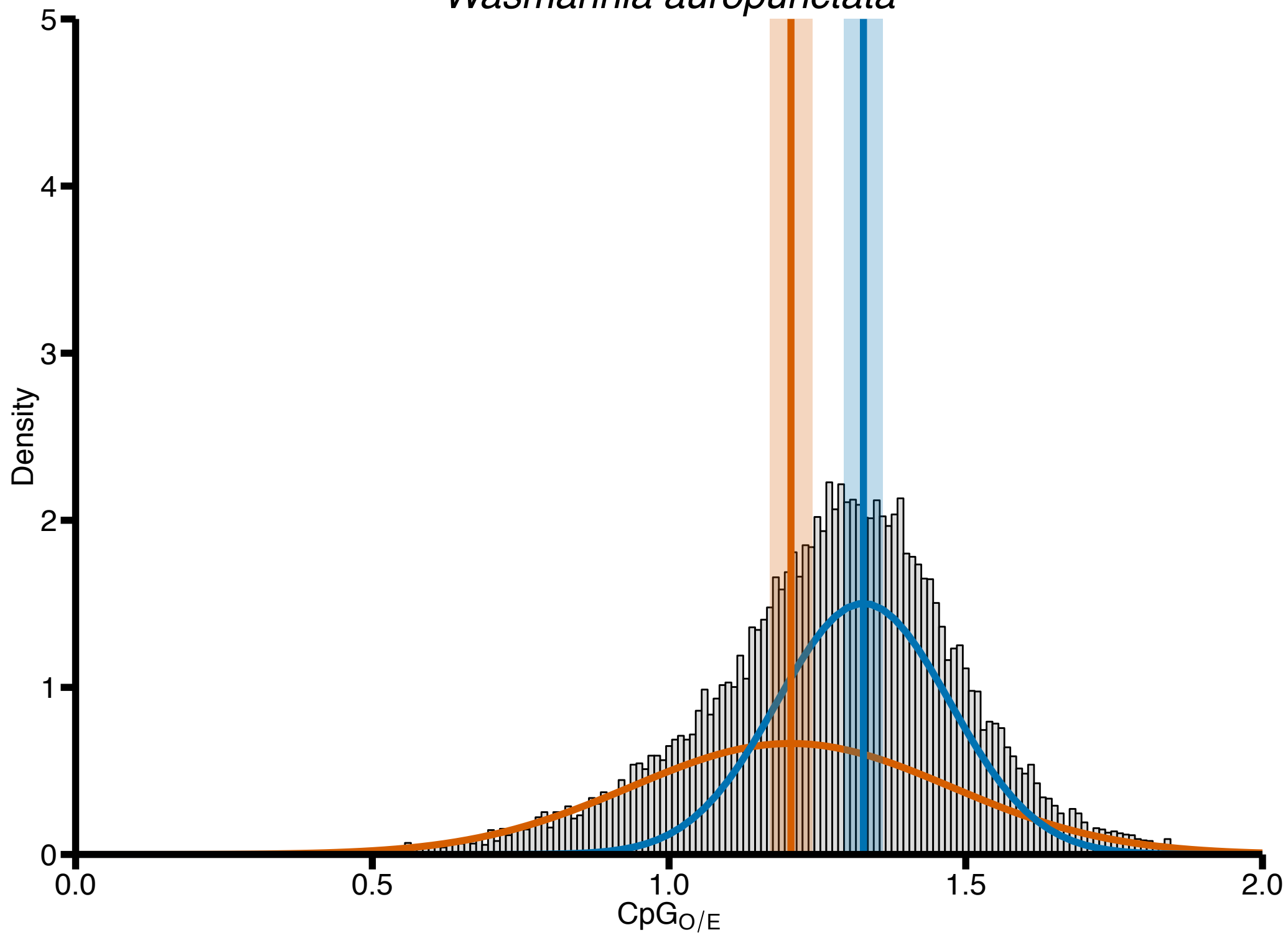
Zootermopsis nevadensis

Amyelois transitella
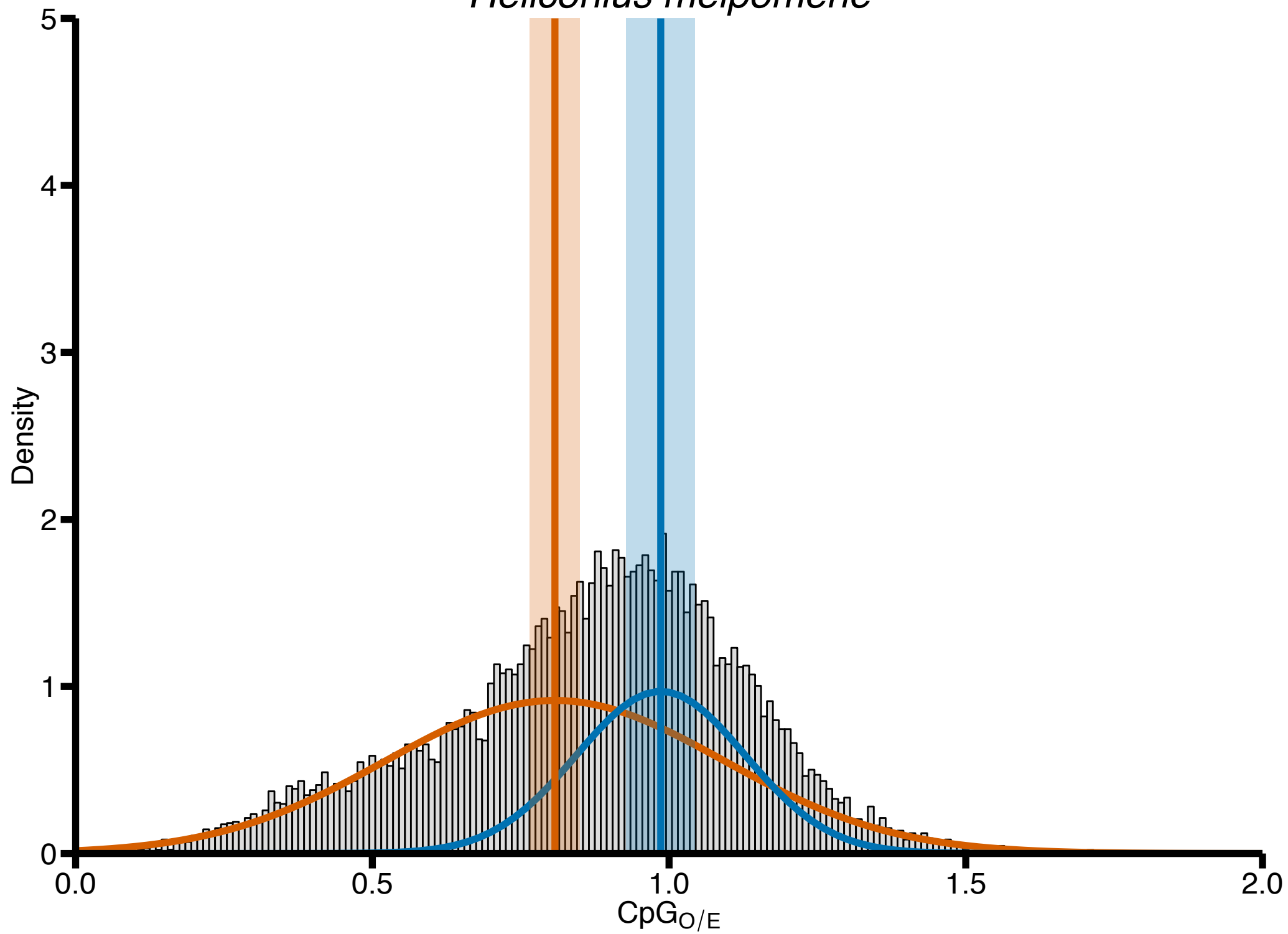
Bombyx mori
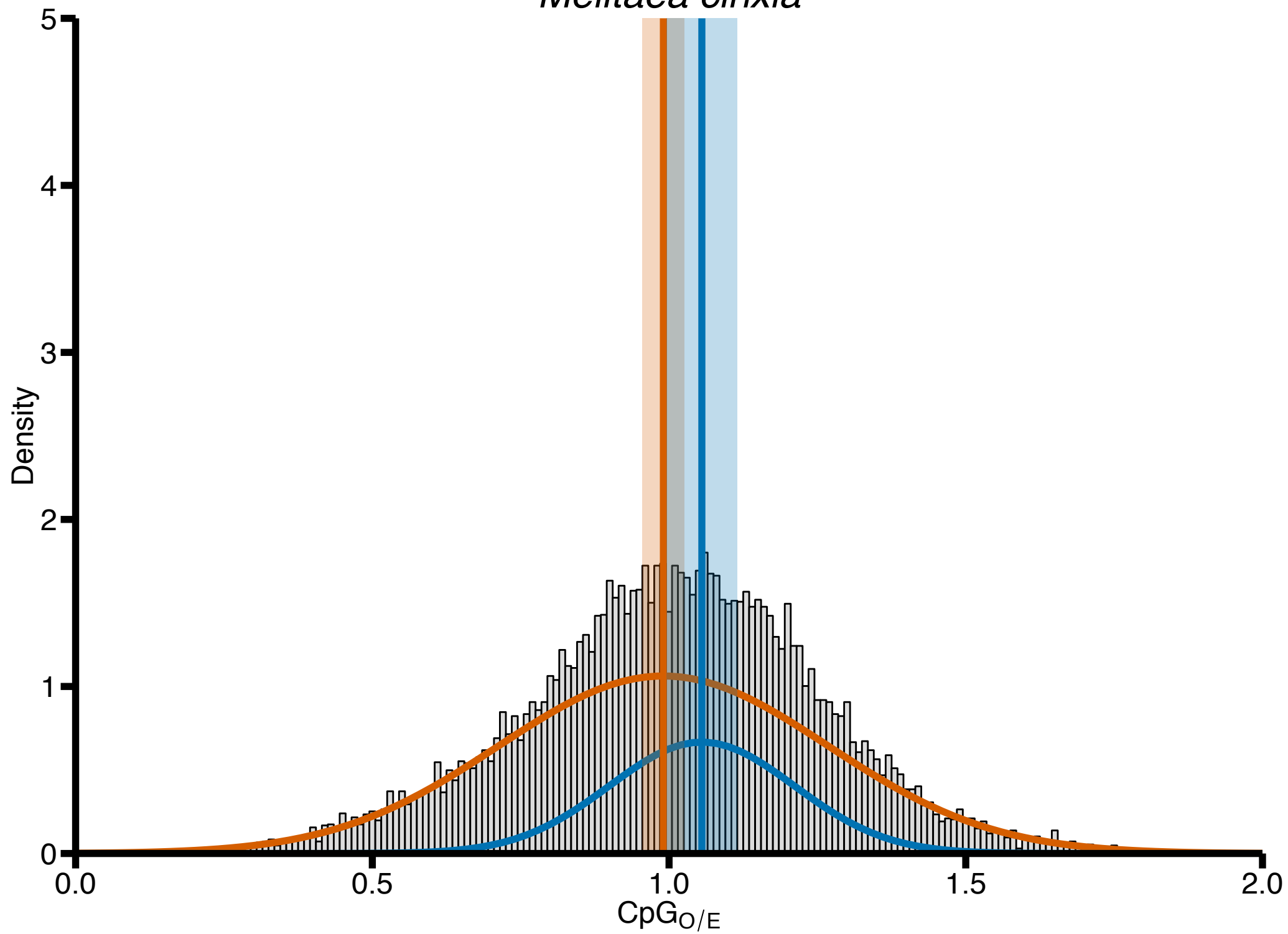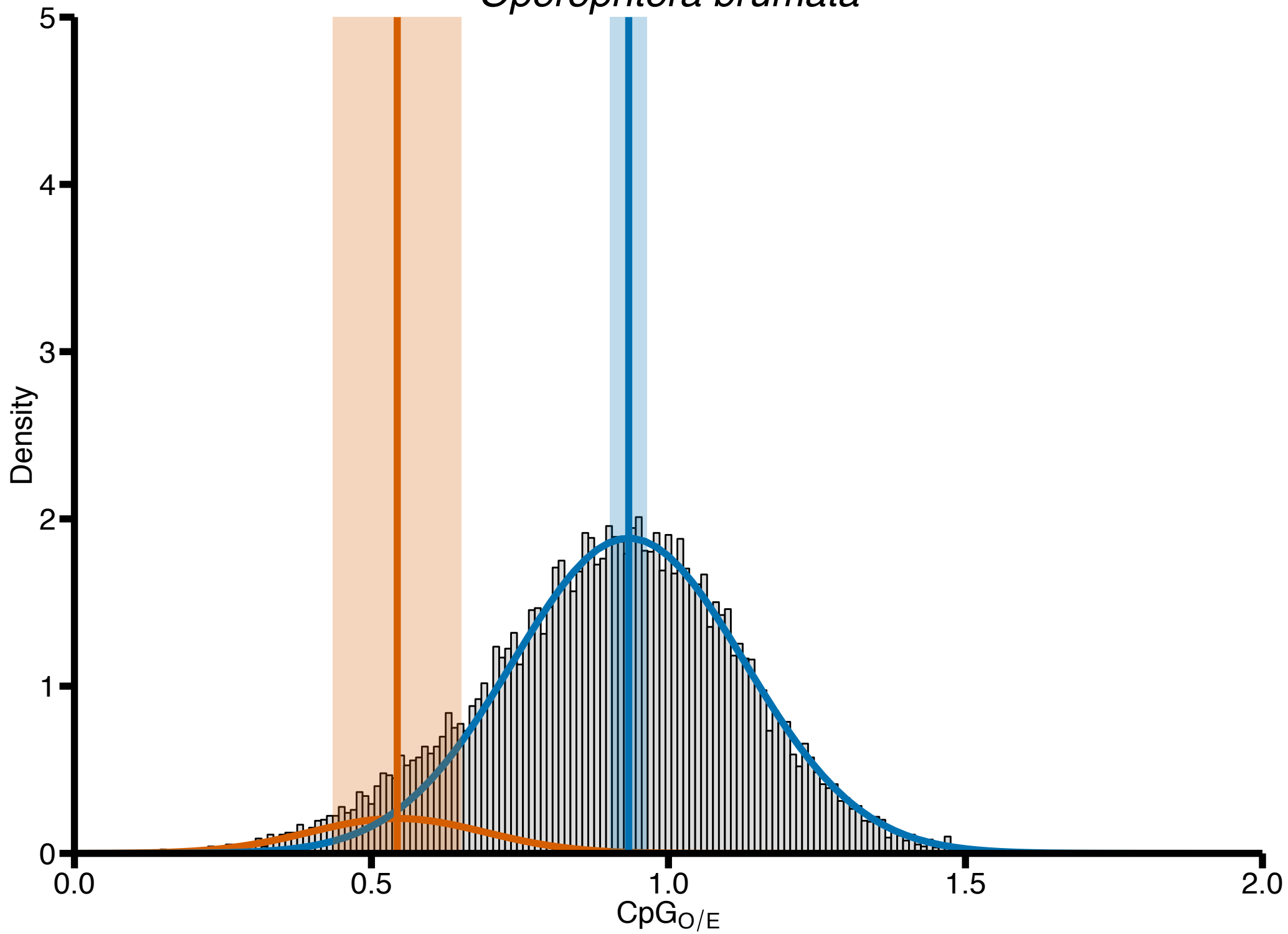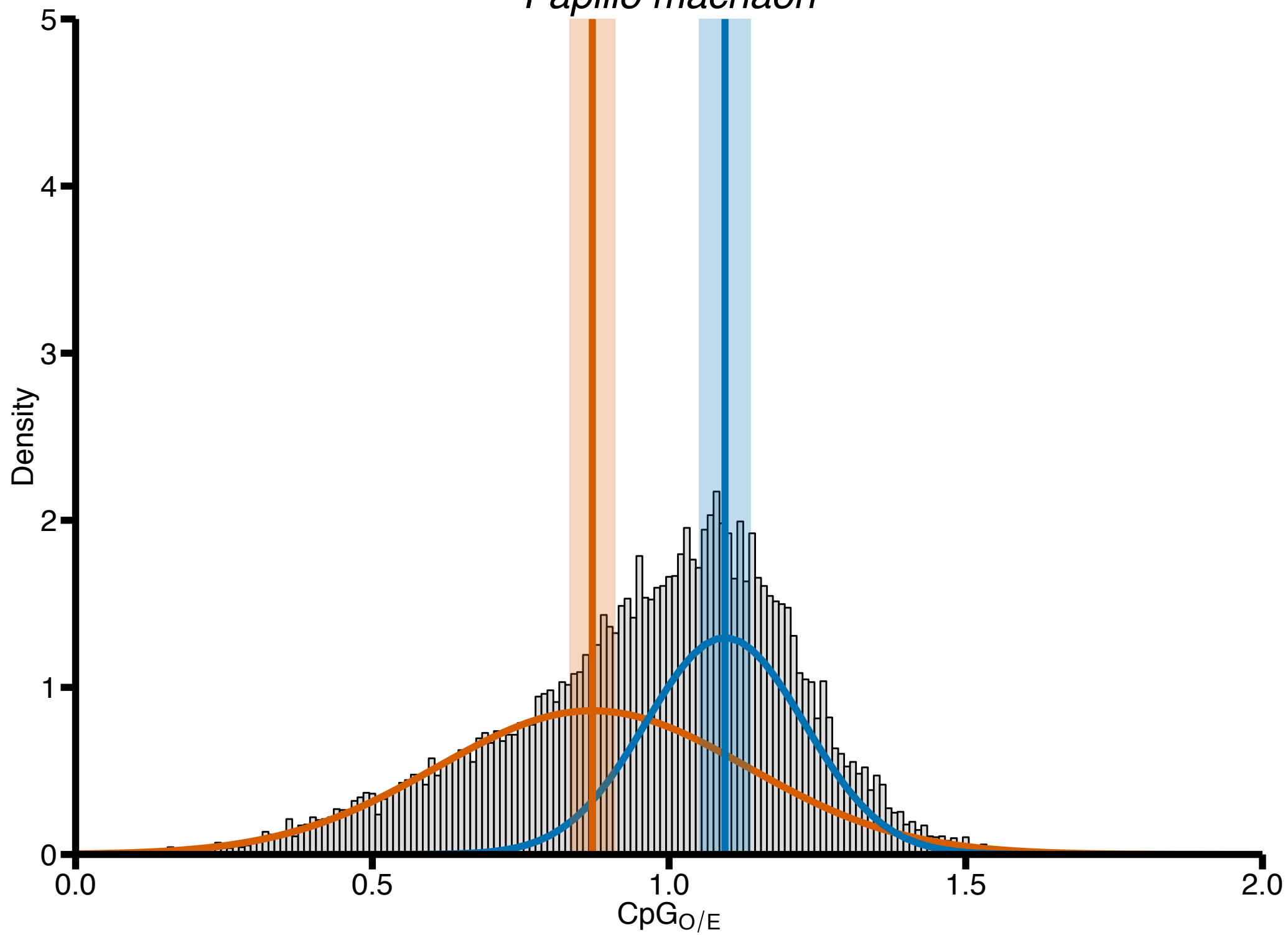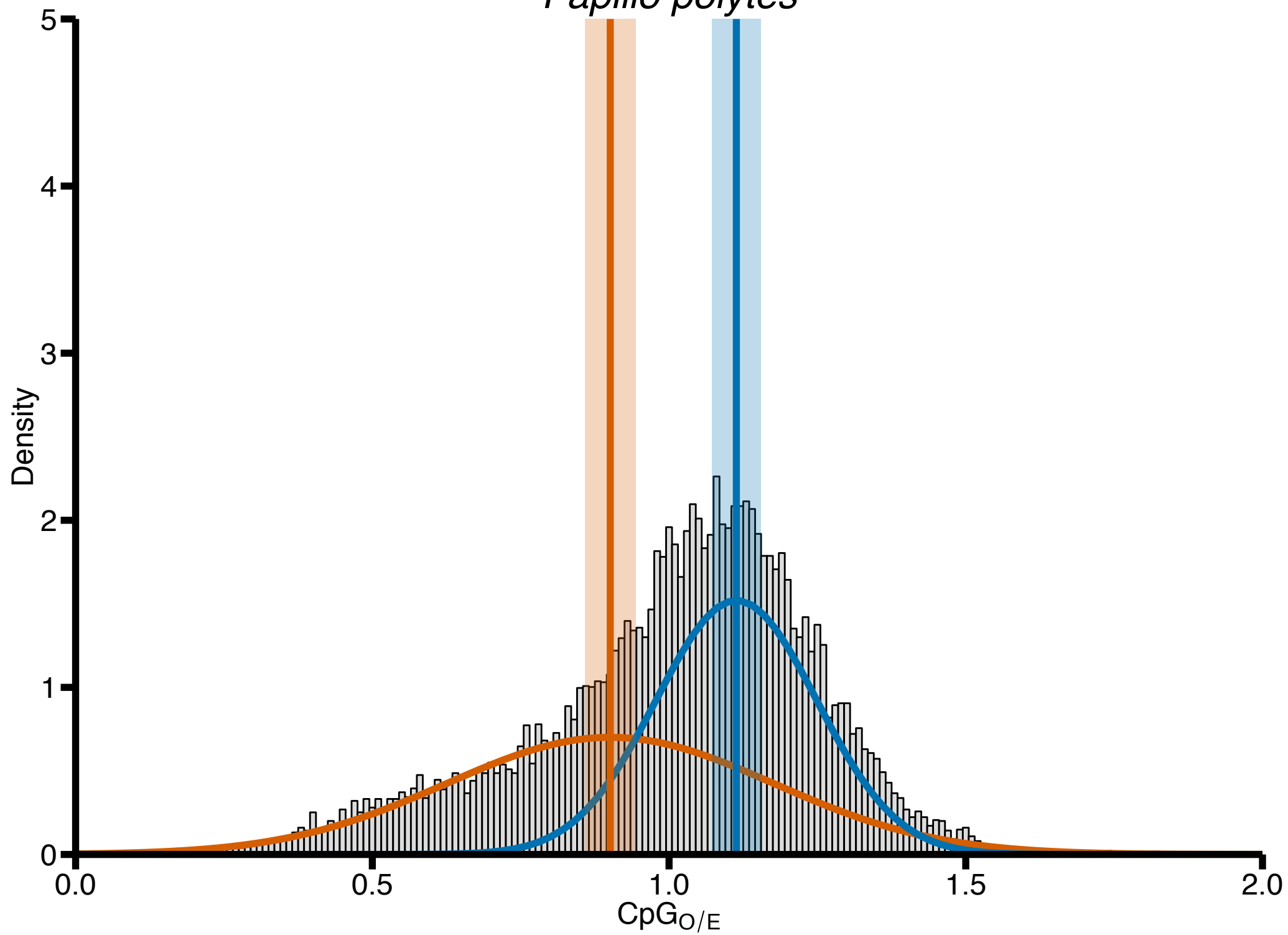
Danaus plexippus

Heliconius melpomene

Melitaea cinxia

Papilio machaon

Papilio polytes

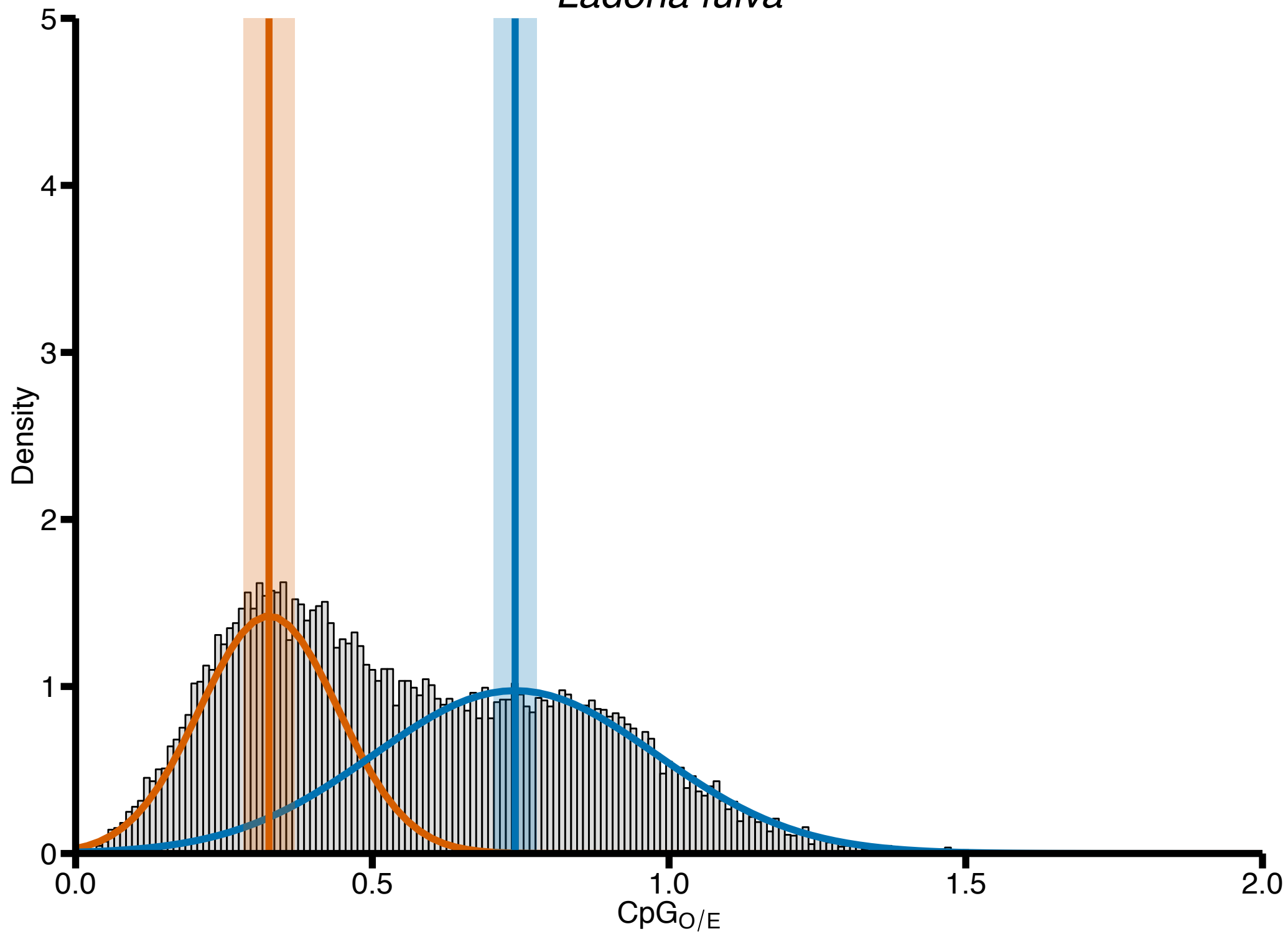*Papilio xuthus*
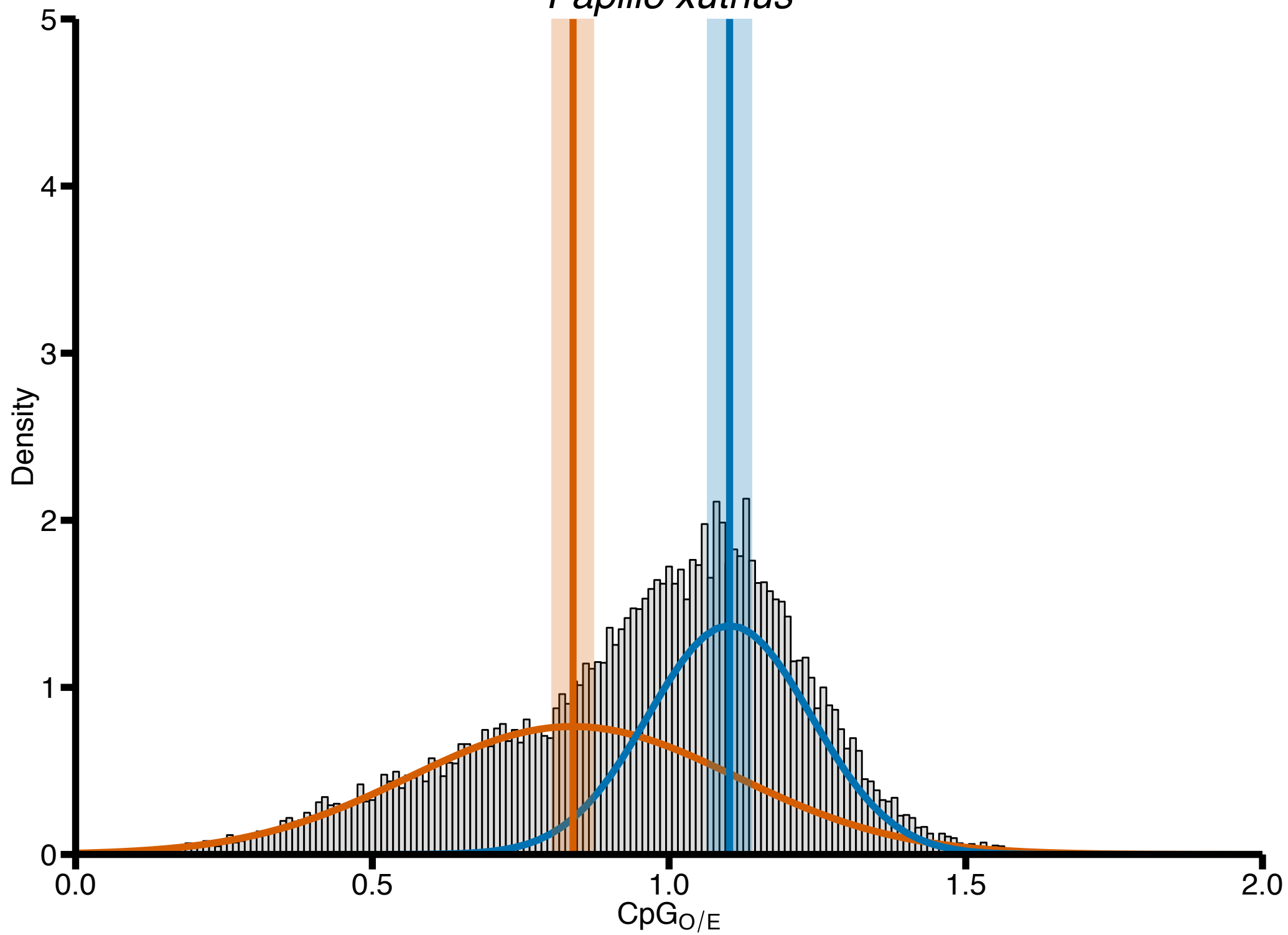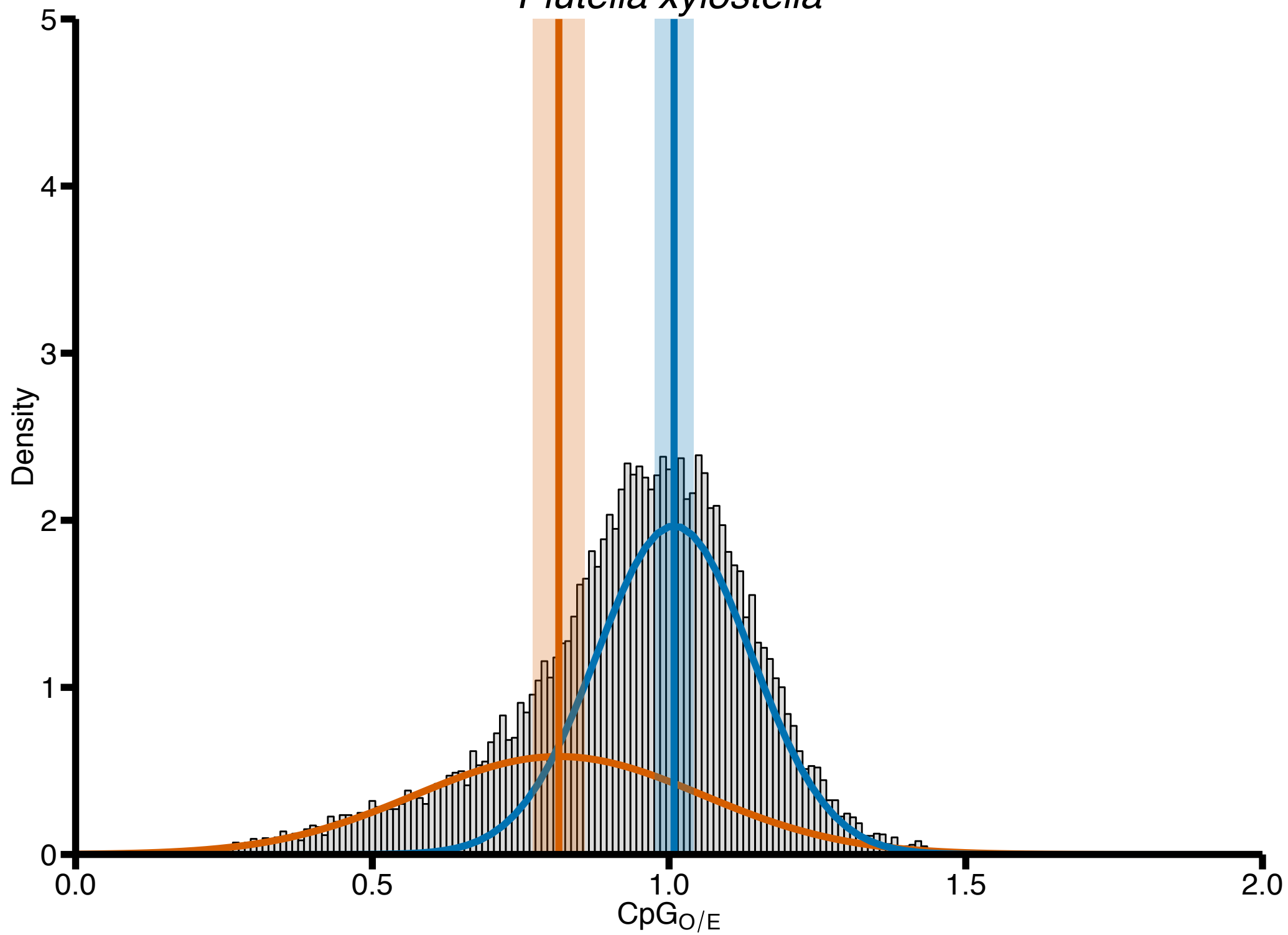
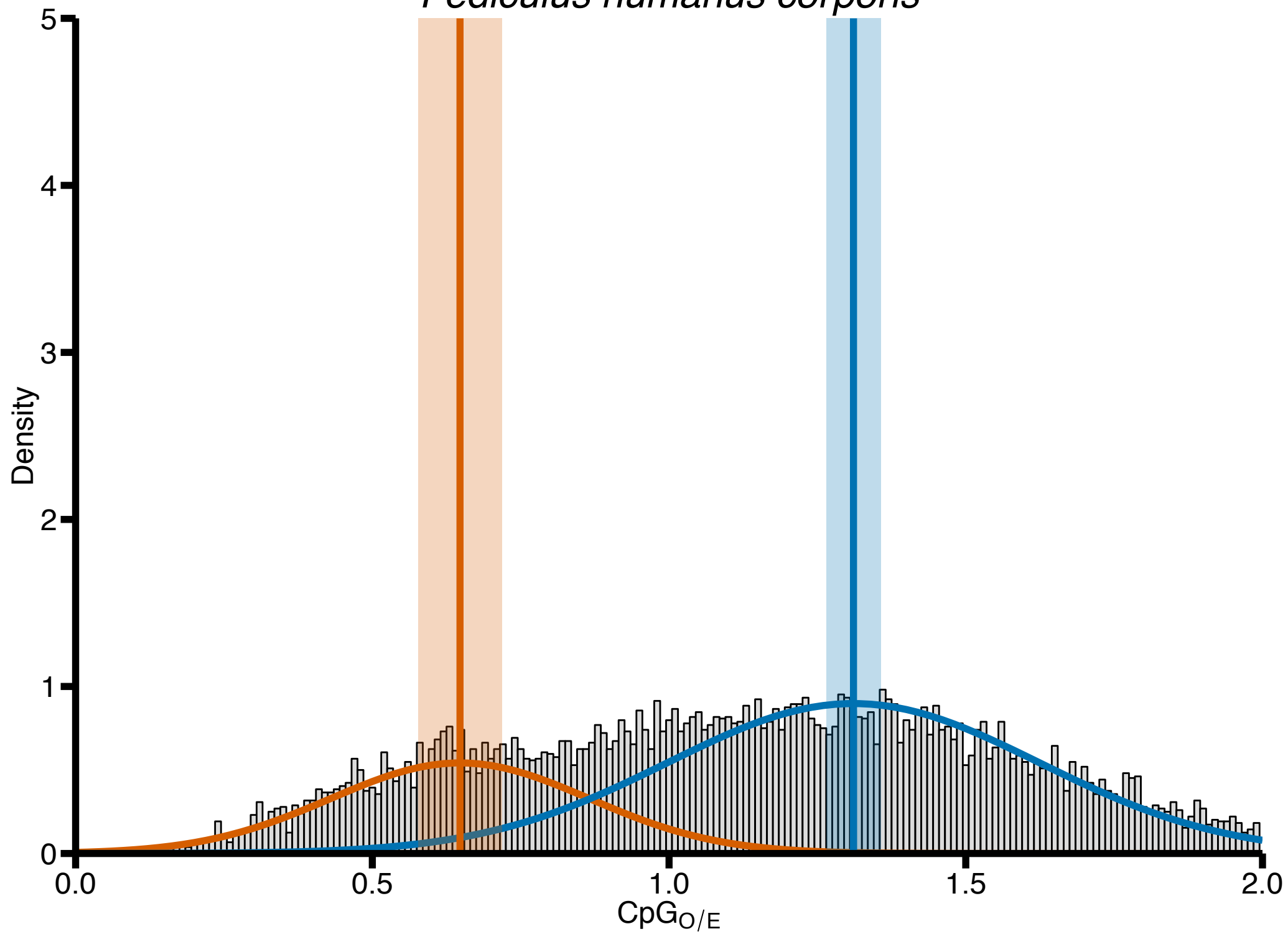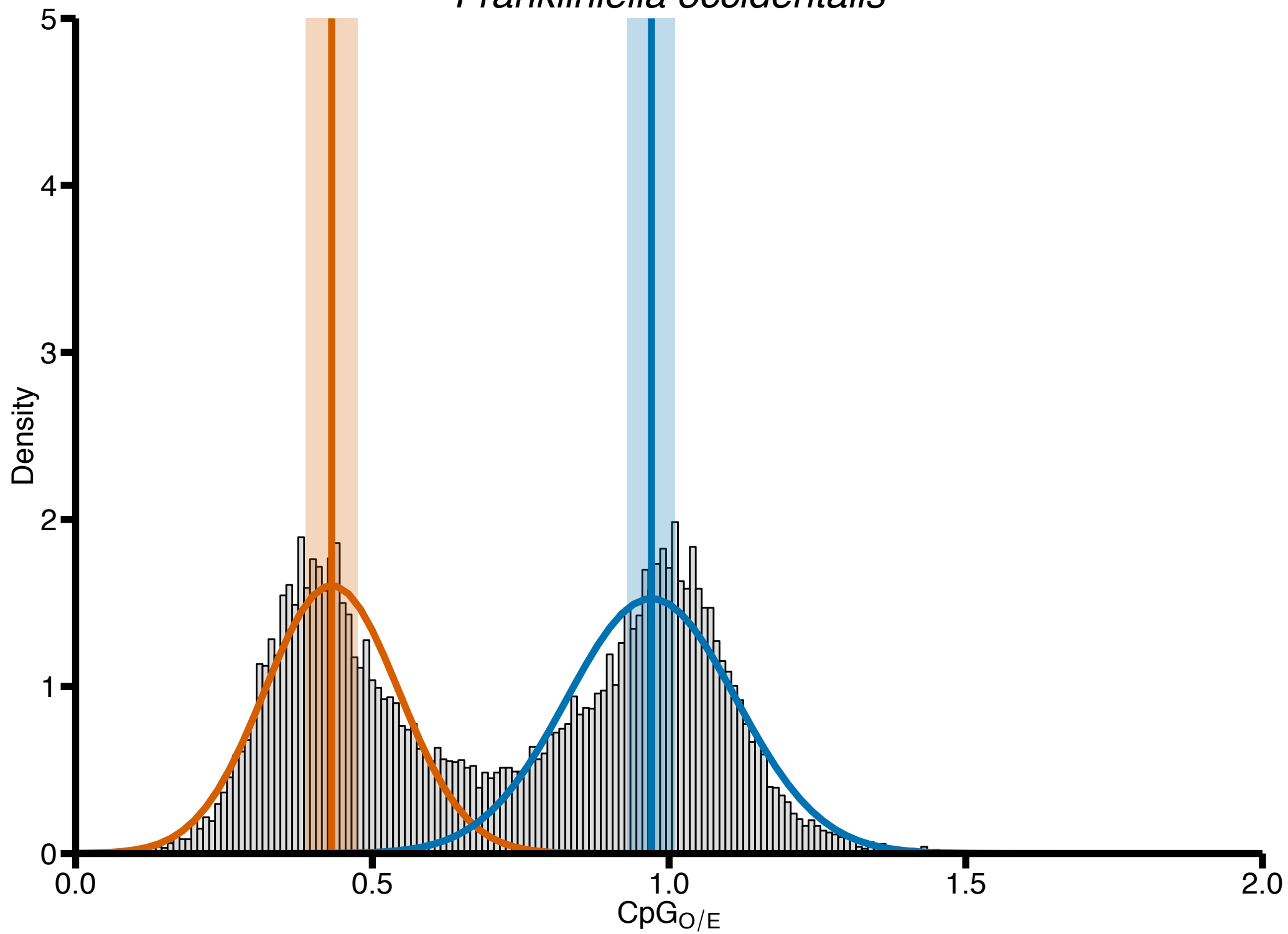*Spodoptera frugiperda*
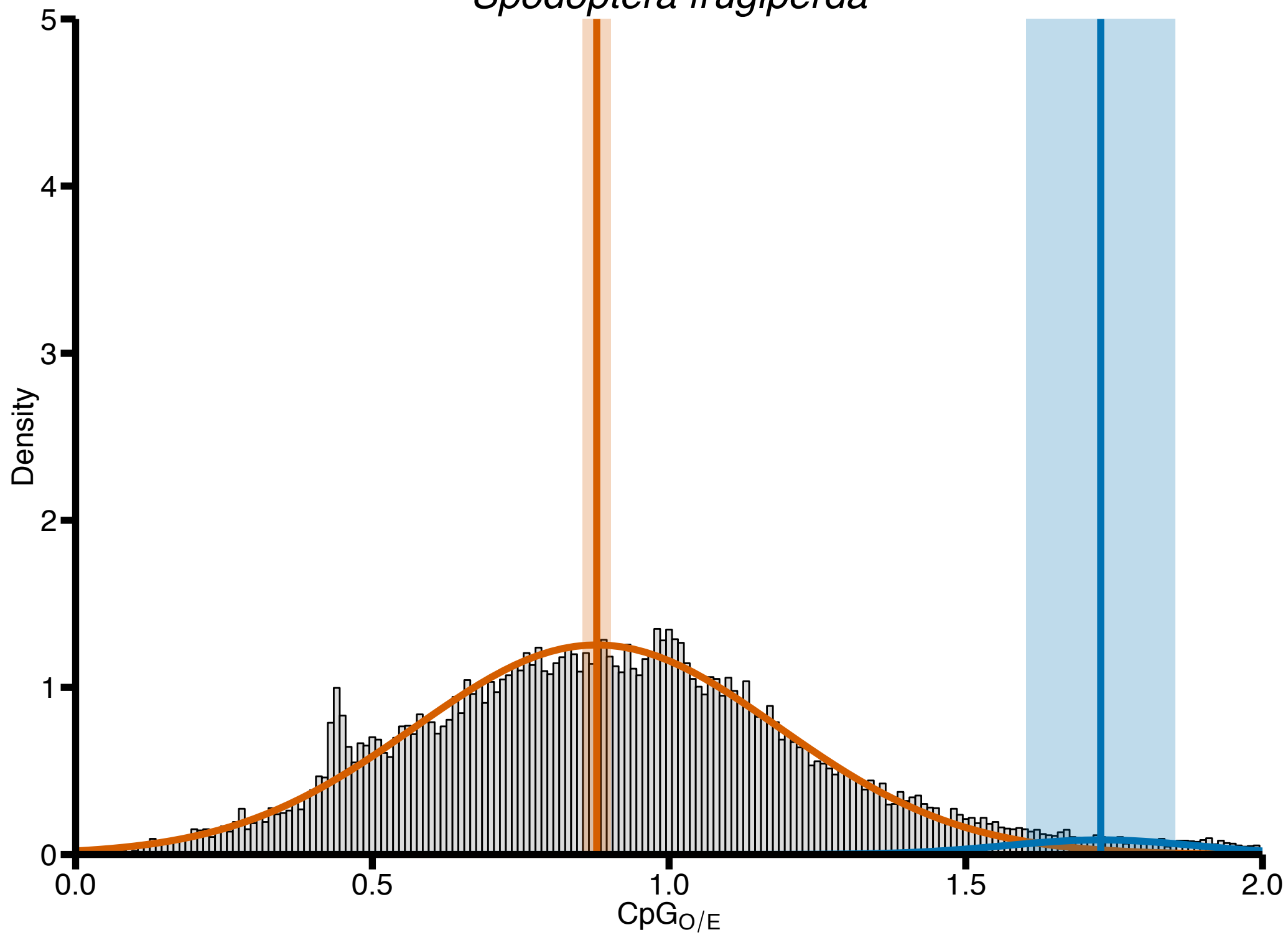
Ladona fulva

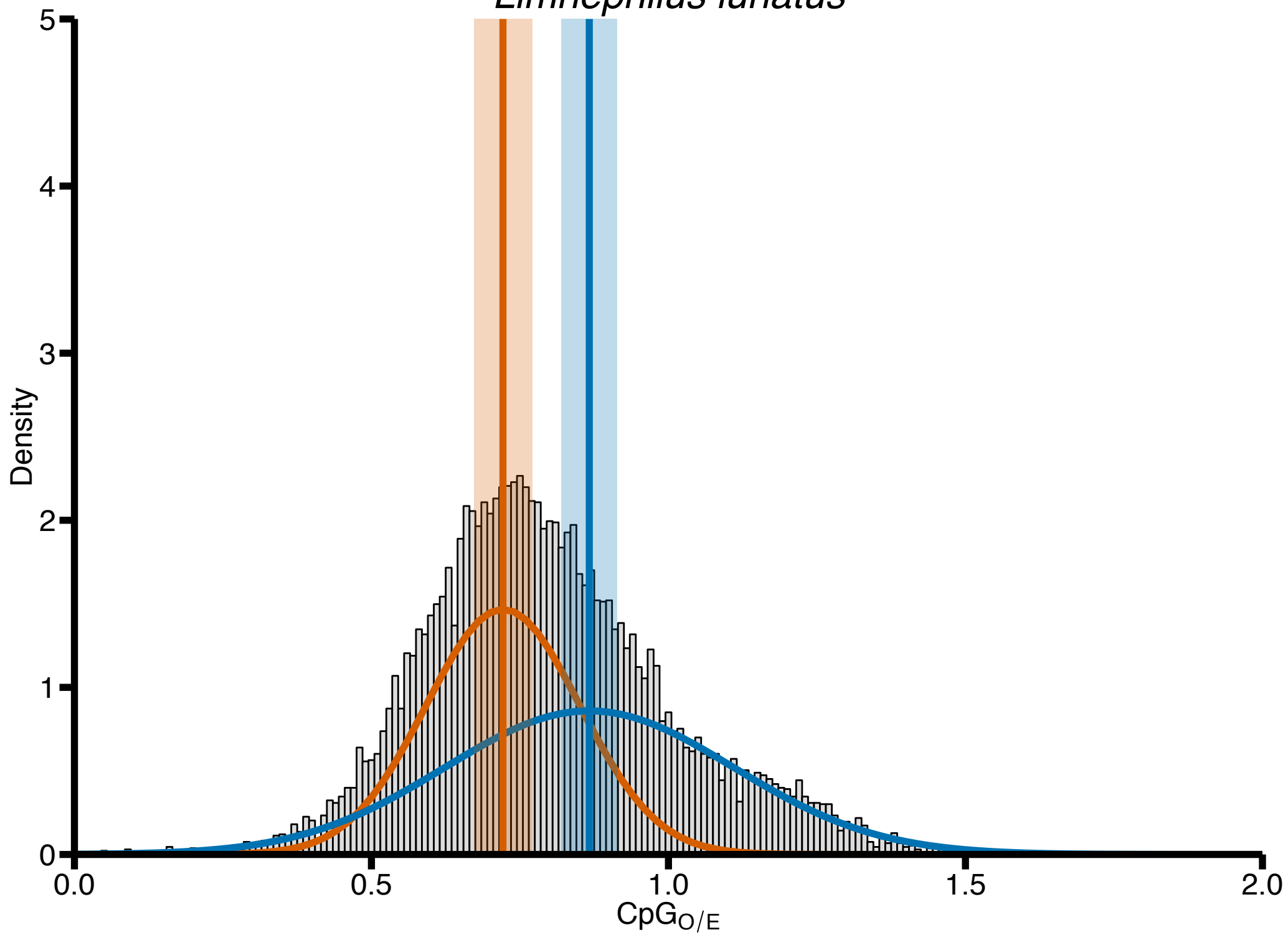*Pediculus humanus corporis*

Frankliniella occidentalis

Limnephilus lunatus

$$y = 3.0578x + 0.005$$
$$R^2 = 0.92444$$

A

Apis dorsata (XP_006612379.1)
Apis mellifera (NP_001185452.1)
Apis mellifera (XP_006569703.1)
Apis florea (XP_012340684.1)
Apis florea (XP_012340685.1)
Eufriesea mexicana (Emex01862)
Bombus terrestris (Bter09047)
Melipona quadrifasciata (Mqua00427)
Bombus impatiens (XP_012246574.1)
Bombus terrestris (Bter09254)
Habropoda laboriosa (Hlab10052)
Lasioglossum albipes (Lalb_11095)
Dufourea novaengliae (Dnov05758)
Megachile rotunda (Mrot03587)
Atta cephalotes (XP_012060955.1)
Acromyrmex echinatior (XP_011056477.1)
Pogonomyrmex barbatus (XP_011631935.1)
Vollenhovia emeryi (XP_011881443.1)
Wasmannia auropunctata (XP_011704997.1)
Monomorium pharaonis (XP_012528413.1)
Monomorium pharaonis (XP_012531701.1)
Solenopsis invicta (XP_011162885.1)
Cerapachys biroi (XP_011350631.1)
Cerapachys biroi (XP_011350635.1)
Camponotus floridanus (XP_011252718.1)
Linepithema humile (XP_012216592.1)
Linepithema humile (XP_012216593.1)
Camponotus floridanus (XP_011257755.1)
Harpegnathos saltator (XP_011151720.1)
Bombus terrestris (Bter05377)
Bombus impatiens (XP_003489082.1)
Melipona quadrifasciata (Mqua05902)
Apis florea (XP_012339009.1)
Apis dorsata (XP_006613898.1)
Apis mellifera (XP_006562865.1)
Eufriesea mexicana (Emex01796)
Habropoda laboriosa (Hlab07979)
Megachile rotunda (Mrot09815)
Dufourea novaeangliae (Dnov00250)
Lasioglossum albipes (Lalb_10913)
Polistes canadensis (XP_014605619.1)
Athalia rosae (XP_012254092.1)
Athalia rosae (XP_012254091.1)
Orussus abietinus (XP_012273739.1)
Orussus abietinus (XP_012287015.1)
Nasonia vitripennis (NP_001164521.1)
Nasonia vitripennis (XP_008212391.1)
Ceratosolen solmsi marchali (XP_011506221.1)
Copidosoma floridanum (XP_014214364.1)
Trichogramma pretiosum (XP_014233345.1)
Nasonia vitripennis (XM_008219729.1)
Nasonia vitripennis (XP_008217946.1)
Ceratosolen solmsi marchali (XP_011505760.1)
Nasonia vitripennis (XM_001607286.2)
Nasonia vitripennis (XP_008202943.1)
Copidosoma floridanum (XP_014203900.1)
Trichogramma pretiosum (XP_014236147.1)
Microplitis demolitor (XP_008547166.1)
Bombyx mori (NP_001036980.1)

99
69
70
94
86
86
99

0.2

Apoidea
Formicidae

DNMT1a
DNMT1b

B

DNMT1a
DNMT1b

Formicidae
Vespidae
Apoidea
Apoidea
Vespidae
Formicidae
Braconidae
Chalcidoidea
Orussidae
Tenthredinidae

☆ Duplication
† Deletion