Journal of Oil Palm Research Vol. 37 (4) December 2025 p. 635-645
DOI: https://doi.org/10.21894/jopr.2024.0039

GENOME SURVEY AND PRELIMINARY GENOMIC CHARACTERISATION OF A KEY OIL PALM POLLINATOR, *Elaeidobius kamerunicus*

NABEEL ATA[1]; MOHD AMIN AB HALIM[1]; MUHAMMAD NURUL YAQIN SYARIF[1]; SAHARUL ABILLAH MOHAMAD[1]; MUHAMMAD DZULHELMI MUHAMMAD NASIR[2]; SIEW-ENG OOI[1] and MEILINA ONG-ABDULLAH[1]*

ABSTRACT

Elaeidobius kamerunicus *is essentially the sole insect pollinator that efficiently pollinates oil palm, a major commodity crop accounting for 31.4% of the world's oils and fats production. Despite its importance, in-depth molecular studies of this pollinator are severely lacking. The scarcity of molecular information to complement the current biological knowledge of* E. kamerunicus *warrants an investigation into the genome features of the pollinator. This aids in understanding the molecular mechanisms of plant-insect interaction during pollination events. Genome size estimation was conducted on Malaysian-bred males and females* E. kamerunicus *using flow cytometry and k-mer analysis on genome survey sequencing data. The cytometric analysis underestimated the genome size compared to the computational sequence-based method with differences of 157-179 megabase (Mb) for the female and male* E. kamerunicus*, respectively. Genome sizes estimated from k-mer analysis were 365.93 Mb (female) and 380.73 Mb (male). Low sequence repeats ratios of 43.49% (female) and 43.70% (male) and high heterozygosity ratios of 1.92% (female) and 1.62% (male) were obtained. The average guanine-cytosine (GC) content of the female genome was 33.48%, and 31.71% for the male. These results lay the groundwork for* E. kamerunicus *genomic studies.*

Keywords: flow cytometry, genome size, genome survey, k-mer analysis, pollinating agent, sex difference.

Received: 26 October 2023; **Accepted**: 16 June 2024; **Published online:** 21 August 2024.

INTRODUCTION

Replacing rubber as the primary economic crop, the oil palm industry is now a major contributor to Malaysia's economy, making oil palm one of the world's leading oil crops (Parveez et al., 2021). In 2021, Malaysian palm oil and its products generated RM108.52 billion in export revenue (Parveez et al., 2022). The global demand for vegetable oils may reach 240 million tonnes by 2050, as estimated by Barcelos et al. (2015). The growing demand for palm oil products, both edible and non-edible, necessitates an increase in yield productivity which has remained at around 3 t ha^{-1} yr^{-1} for decades (Woittiez et al., 2017). To support sustainability of the oil palm industry, increasing production without expanding into new land for oil palm cultivation is much needed. Enhancing in-field palm fruit formation can potentially improve oil yield to meet the growing demand for palm oil.

Proper fruit development therefore requires efficient pollinators, usually insects, for the fertilisation of flowers. The oil palm (*Elaeis guineensis*) is a monoecious plant, where independent male and female flowers developed separately on the same plant. *Elaeidobius kamerunicus*, an entomophilous pollinator, is specific to the oil palm, with male flowers serving

[1] Malaysian Palm Oil Board,
6, Persiaran Institusi, Bandar Baru Bangi,
43000 Kajang, Selangor, Malaysia.

[2] Crop Protection & Bio-Solutions,
FGV R&D Sdn. Bhd.,
Tun Razak Agricultural Research Centre,
27000 Jerantut, Pahang, Malaysia.

* *Corresponding author e-mail: meilina@mpob.gov.my*

as their preferred site for feeding, laying eggs and reproducing (Meléndez & Ponce, 2016). Both male and female weevils are vital for effective pollination, despite slight body size differences. Their pollen-carrying abilities are comparable (Dzulhelmi et al., 2022). Both are attracted by the anise-like scent of estragole released by male and female inflorescences during anthesis, enabling pollen transfer from the weevil to the stigma of female inflorescences. Its introduction into Malaysia in the early 1980s has had a positive impact on the oil palm industry, significantly enhancing palm oil production (Mohamad et al., 2023). Malaysia's success paved the way for the global introduction of *E. kamerunicus* in other oil palm-growing nations like Indonesia, Papua New Guinea, Ecuador, Colombia, Costa Rica and Central and South America. This has notably boosted oil yields in these countries (Appiah & Agyei Dwarko, 2013). This underscores the economic significance of the weevil to the palm oil industry, highlighting the importance of safeguarding their well-being.

Despite being a key oil palm pollinator, little is known about the molecular biology of *E. kamerunicus*. Apriyanto and Tambunan (2021) briefly studied its genome, based on a single male individual. A total of 26,566 predicted genes and 281,668 simple sequence repeat (SSR) markers were annotated on their 269.79 Mb genome assembly. Genome size, or deoxyribonucleic acid (DNA) C-value, is the total DNA in a haploid chromosome set, crucial for evolution, species differentiation, and hybrid identification (Bureš et al., 2004; Morgan-Richards et al., 2004; Zonneveld, 2001). Closely related species can have varying genome sizes due to repetitive sequence changes such as segmental duplications or deletions driven by environmental adaptation (Boulesteix et al., 2006; Biémont, 2008). As of 2021, genome size data exists for only 0.135% of the nearly one million known insect species, with just nine reports on curculionids (Gregory, 2023).

To estimate the genome size of a species, flow cytometric analysis was traditionally seen as the "gold standard" (Mounsey et al., 2012). However, the rapid progress of next-generation sequencing has led to the use of computational sequence-based methods, now common in many insect genome projects (Pflug et al., 2020). As molecular information on *E. kamerunicus* is rather limited, we first set out to determine the genome size of both sexes of *E. kamerunicus* as different sexes of an organism may have different genome sizes. In our study, we employed flow cytometry to estimate the genome sizes of both male and female *E. kamerunicus* weevils. We then expanded our research by using k-mer analysis with Illumina genome skim sequencing data following the procedure of Sarmashghi et al. (2019). We compared the genome size estimates from these two methods. Our results enhanced the comprehension of genome size in the insect taxa, especially within the curculionidae family. This sequencing data also offers a valuable genomic foundation for future functional studies. The data on both sexes, together with the male genome sequence by Apriyanto and Tambunan (2021), can be used to support future efforts to generate a good *E. kamerunicus* reference genome.

MATERIALS AND METHODS

Sampling of Insects

Male and female *E. kamerunicus* weevils were captured at the MPOB Research Station in Jerantut, Pahang, Malaysia (GPS coordinate: 4°17'14"N 102°24'44"E). The weevils were randomly sampled from male inflorescences. Briefly, the fully anthesised spikelets from male inflorescence were excised and pulled into two whereby the weevils were then easily accessible. The male and female *E. kamerunicus* were manually separated (*Figure 1a*), flash-frozen in liquid nitrogen, and stored at -80°C for DNA extraction and flow cytometric analysis. *Drosophila melanogaster* was used as the external reference standard. *Drosophila* cultures were obtained from the Universiti Kebangsaan Malaysia (UKM) and maintained under a 12 hr light: 12 hr dark photoperiod at 25 ± 2°C in the laboratory.

DNA Isolation, Library Construction and SkimSeq Sequencing

DNA was extracted with the DNeasy Blood and Tissue kit (Qiagen) following the manufacturer's instructions. Genomic DNA quantity and quality were assessed with the Biotek Instruments Synergy HTX multi-mode reader, followed by visualisation on a 1% agarose gel. Seven male and seven female *E. kamerunicus* DNA samples were sent to Novogene, USA, for library preparation and whole-genome sequencing using the Illumina HiSeq platform. DNA from seven individuals was adequate for library preparation with the NEBNext® DNA Library Prep Kit (NEB). Briefly, 1,500 nanograms (ng) of genomic DNA was randomly sheared into approximately 350 base pairs (bp) fragments. The sheared DNA was subjected to library construction followed by end repair, dA-tailing, and further ligation with NEBNext adapters. The required fragment sizes (300-500 bp) were PCR-enriched with P5 and indexed P7 oligos. Paired-end sequencing (PE150) was conducted on the libraries. The quality control

steps for the raw sequences included removal of the adapter sequences, discarding paired reads with 10% or more ambiguous nucleotides on either read, and eliminating paired reads if over 50% of either read contained low-quality nucleotides (base quality ≤5) (Cock et al., 2009). The *E. kamerunicus* SkimSeq sequencing data is available at http://genomsawit.mpob.gov.my/index.php?track=30&nu=1&info=3.

K-mer Estimation of Genome Size, Heterozygosity Ratio and Repeated Sequences

K-mer analysis was performed on clean reads from both males and females to predict genome size, assess heterozygosity ratio, and identify repeated sequences before assembly analysis. Typically, k-mer values between 17 and 27 (Pflug et al., 2020), are used for prediction and analysis. The genome size was calculated by dividing the total k-mer count with the highest value in the k-mer frequency distribution, according to the following formula: Genome size = k-mer number/k-mer depth (Marçais & Kingsford, 2011). The revised genome size was then calculated by excluding the k-mer error. The Genomeye program (Novogene) was used to determine the genome size, heterozygosity rate and repeat content.

Flow Cytometry Analysis

Samples were prepared according to Galbraith et al. (1983), with minor modifications. Approximately seven adult weevils and seven *Drosophila* heads were dissected on a petri dish. The samples were fully homogenised in 700 µL ice-cold Galbraith's buffer (pH 7.0) containing 45 mM $MgCl_2$, 20 mM MOPS [3-(N-morpholino) propanesulfonic acid], 30 mM sodium citrate and 0.1% Triton X-100. The homogenate was filtered into a 1.5 mL tube through a 40 µm cell strainer and incubated with RNase at a final concentration of 20 µg mL^{-1} for 10 minutes at 25°C. The solution was then centrifuged at 8,000 rpm for 5 min. The pellet was resuspended in 700 µL phosphate buffer (pH 7.0), stained with 50 µL propidium iodide (PI) (1 mg mL^{-1}) and incubated in darkness at 4ºC for 10 min. The suspension obtained in the final step was analysed using the BD FACSCalibur™ flowcytometry system (BD Biosciences). Cellular DNA content was measured using the fluorescence intensity of each sample exposed to 488 nm wavelength. The same parameter settings were applied to *D. melanogaster* head samples. Estimation of the nuclear DNA content was based on *D. melanogaster*, 1C = 175 Mb, as an external reference standard. Fluorescence intensity from all samples were compared with the fluorescence intensity of *D. melanogaster* to obtain the ratio, then multiplied by the genome size of *D. melanogaster* (1C = 175 Mb). The fluorescence intensity peaks and genome size of samples were analysed using CellQuest3^R software (BD Biosciences).

The total quantity of DNA in each sample was calculated as Equation (1):

$$\text{Sample IC value} = \frac{\text{Sample 2C mean peak position}}{\text{Reference 2C mean peak position} \times \text{IC reference}} \quad (1)$$

The genome size (Mb) is reported as 1C, i.e., the mean amount of DNA in one copy of a single complete genome. One megabase is equal to one million base pairs (1 Mb = 1,000,000 bp). The entire flow cytometry experiment was repeated for a total of five replications, amounting to the use of a total of 35 males and 35 females *E. kamerunicus* individuals, and 70 *D. melanogaster* individuals.

***De novo* Genome Assembly**

The SOAPdenovo assembler (Li et al., 2010) was used with a k-mer of 41 and default parameters to create an initial genome assembly based on a de Bruijn graph (Luo et al., 2012). After that, GC content and average sequencing depth were calculated to assess sequencing bias. To assess data contamination, 10,000 pairs of processed reads were randomly selected for a BLASTN search against the National Center for Biotechnology Information (NCBI) nonredundant (NR) nucleotide database with defined options for processors, program name, and expectation value (-a 6 -p blastn -e 1e-05). Only sequences above 100 bp were selected for assembly statistics.

RESULTS AND DISCUSSION

Genome Size Prediction of *E. kamerunicus*

This study aimed to estimate the genome sizes of both the male and female *E. kamerunicus* (Coleoptera; Curculionidae) and to also supplement the recently reported male genome size (Apriyanto & Tambunan, 2021). Genome size estimation was carried out with two approaches, a k-mer sequence-based approach and flow cytometric analysis. A single major peak in the k-mer plot provides a good approximation of the actual genome size (*Figure 1*). Thus, the estimated genome sizes through k-mer analysis for *E. kamerunicus* were 365.93 Mb for the female and 380.73 Mb for the male (*Table 1*), suggesting that the male weevil genome is bigger than the female.

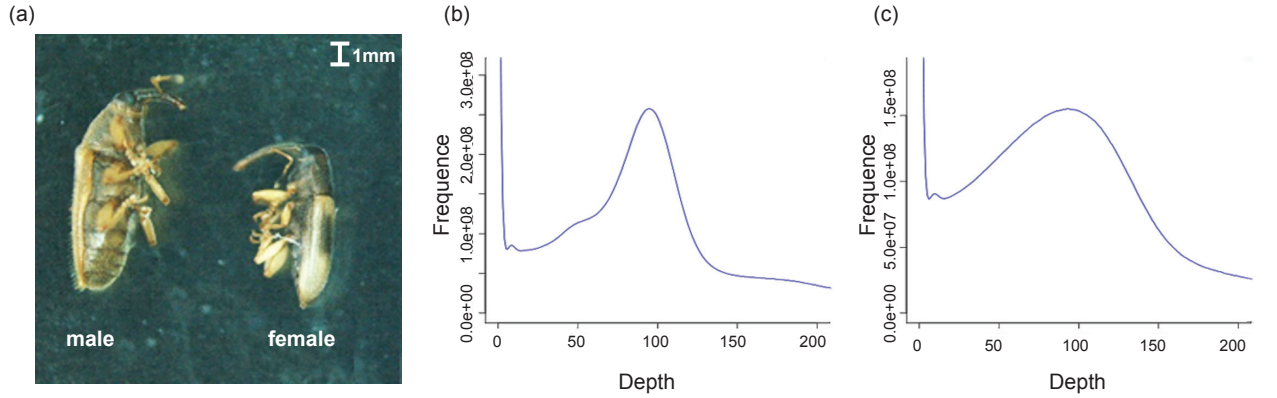

Figure 1. E. kamerunicus *morphology and k-mer analysis plots. (a) Male and female* E. kamerunicus *adult weevils. Distribution of 17-mer depth and frequency for (b) female, and (c) male* E. kamerunicus. *The x-axis shows k-mer depth, and the y-axis represents the proportion relative to the total frequency.*

TABLE 1. K-MER ANALYSIS AND *E. kamerunicus* GENOME SIZE EVALUATION, COMPARED TO GENOME SIZE ESTIMATED THROUGH FLOW CYTOMETRY

Sex	k-mer depth	Total number of 17-mer	Estimated genome size from k-mer analysis (Mb)	Heterozygosity rate (%)	Repeat rate (%)	Revised estimated genome size from k-mer analysis (Mb)	Estimated genome size (haploid) from flow cytometry (Mb)
Female	92	3.43×10^{10}	372.53	1.92	43.49	365.93	208.74
Male	94	3.64×10^{10}	386.75	1.62	43.70	380.73	201.88

Using flow cytometry analysis, the *E. kamerunicus* genome was estimated to be slightly larger than *D. melanogaster* based on the differences in their fluorescence intensities (*Figure 2*). The DNA content was converted into genome size using the relationship: 1 picogram (pg) of DNA equals 980 Mb (Bennett et al., 2000). The mean nuclear DNA content of *E. kamerunicus* was 0.213 ± 0.014 pg for female (n = 5) and 0.206 ± 0.002 pg for male (n = 5) (*Table 2*). Using the known fruit fly genome size of 175 Mb, we estimated the haploid *E. kamerunicus* genome sizes at approximately 208.74 Mb (female) and 201.88 Mb (male) (*Table 1*). Flow cytometry analysis suggested that the male genome was smaller than the female genome, contrary to the k-mer analysis results. Furthermore, the genome size estimates from k-mer analysis were approximately twice the size from flow cytometry estimates.

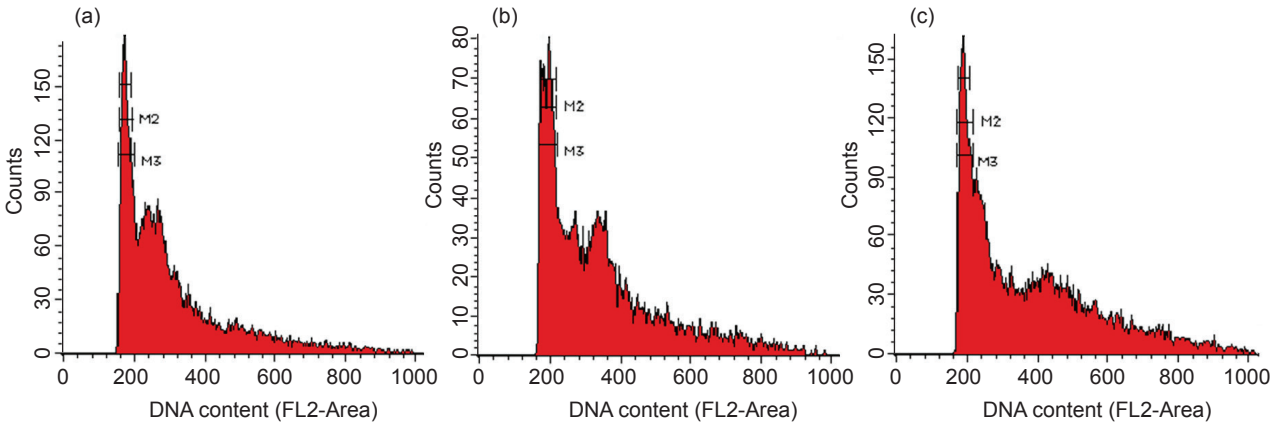

Figure 2. Flow cytometry profiles of (a) D. melanogaster *reference standard, (b) female, and (c) male* E. kamerunicus.

TABLE 2. FLOW CYTOMETRIC ESTIMATION OF GENOME SIZE USING THE MEAN NUCLEAR DNA CONTENT

Male/replicate	DNA content/1C value (pg)
1	0.206
2	0.207
3	0.205
4	0.204
5	0.206
Mean ± SD	**0.206 ± 0.002**
Femalereplicate	**DNA content/1C value (pg)**
1	0.214
2	0.204
3	0.221
4	0.221
5	0.206
Mean ± SD	**0.213 ± 0.014**

Note: pg - picogram; SD - standard deviation.

Differences of 6.86 and 14.80 Mb between the fluorescence and sequence-based analyses were obtained for the male and female genome sizes, respectively. The 13.72 Mb difference between male and female diploid genome sizes estimated through fluorescence-based analysis is relatively small, roughly 3.4% of the estimated male genome. The 14.8 Mb difference between the sexes, estimated through sequence-based analysis, also accounts for only approximately 3.9% of the estimated male genome size. Both methods yielded conflicting estimates: Cytometric analysis suggested a larger female genome, while k-mer analysis indicated a larger male genome. Therefore, the k-mer approach overestimated the male genome by about 178.85 Mb and the female genome by about 157.19 Mb compared to the cytometric technique. Apriyanto and Tambunan (2021) suggested that the *E. kamerunicus* has a genome size of 269.79 Mb (based on a single male), which was larger than our cytometry estimate by 69.90 Mb but smaller by 110.94 Mb when compared to k-mer estimates. The use of pooled weevil individuals, rather than the single isolated male in Apriyanto and Tambunan (2021), may be a reason for the genome size differences. However, pooling of several individuals was necessary to provide sufficient DNA for the SkimSeq technology. Genome size estimates using flow cytometry and sequence-based analyses for other insects also showed similar discrepancies. The genome size of *Bemisia tabaci* was ~682.30 Mb based on k-mer analysis and ~690.00 Mb according to flow cytometric analysis, using chicken red blood cell as the standard reference. This represents just a 10.00 Mb difference (Chen et al., 2015). Conversely, Guo et al. (2015) reported a larger genome size for *B. tabaci* using the k-mer approach (720 Mb), which is ~60 Mb larger than the ~653 Mb estimate from flow cytometry, with *D. melanogaster* as the reference. These studies highlight that, apart from the choice of reference standards, cytometric analysis often provides variable estimates compared to k-mer methods. Other comparative studies, such as He et al. (2016), found differences in genome size estimates using k-mer and cytometric analyses. For *Laodelphax striatellus*, estimates were 657 and ~555 Mb, a 102 Mb difference. In the case of *Micropentila cingulum*, k-mer analysis underestimated the genome size at 136 Mb, 25 Mb less than the ~157 Mb from cytometric analysis. Pflug et al. (2020) concluded that although certain sequence-based methods, like k-mer distribution or read mapping, aligned with flow cytometry estimates for some species, no single technique consistently agreed with flow cytometry. While these two methods are expected to yield similar results, there are often some inconsistencies. These variations may be due to the use of different external standards, tissue samples at various growth stages in flow cytometry and potential cell endoreplication in insects (Yu et al., 2021).

The average coleopteran species (Order) genome size is 760 Mb (oscillating from 160- 5,020 Mb (Gregory, 2023). On the other hand, the average genome size for the curculionids species (Family) is 568 Mb (i.e., 205-842 Mb) (Gregory, 2023), though only 10 species were recorded, with the biggest being the rice water weevil, *Lissorhoptrus oryzophilus* (981 Mb). Holometabolous insects, like *E. kamerunicus*, which undergo complete metamorphosis with egg, larvae, pupae and adult stages, have been hypothesised to be constrained to a genome size limit of 2 pg or 956 Mb (Gregory, 2002; Hanrahan & Johnston, 2011). Significant gaps in taxon sampling for holometabolous insects persist, with entire orders still unrepresented. It is unclear whether this perceived restriction signifies an actual genome size threshold for holometabolous insects, or if expanded sampling might reveal larger genomes among them. Still, the estimated *E. kamerunicus* genome size from this study remains below the size limit for curculionids, which supports the hypothesis mentioned earlier. Our findings, using k-mer and flow cytometric analyses, showed that *E. kamerunicus* has a haploid genome size (C-value) of between 201-380 Mb (0.20-0.38 pg). Specifically, the genome size of the male *E. kamerunicus* is 202-381 Mb and 209-366 Mb for the female *E. kamerunicus*. The male genome size of 269.79 Mb reported by Apriyanto and Tambunan (2021) is within the range tabulated from this study.

Regarding the genome sizes estimated through genome sequencing assembly, beetle genomes have generally been on the smaller side, with an average assembly size of 286 Mb, ranging from 160 to

782 Mb (Cunningham et al., 2015; Hazzouri et al., 2020; Keeling et al., 2013; Meyer et al., 2016; Parisot et al., 2021; Powell et al., 2021; Richards et al., 2008; Vega et al., 2015). The model species for beetles, *Tribolium castaneum*, has a relatively small genome size of 160 Mb only (Richards et al., 2008) with a later improvement to 165 Mb (Herndon et al., 2020). The first assembled curculionidae genome, *Dendroctonus ponderosae*, is 205.8 Mb (http://www.genomesize.com), which is relatively similar to the 204 Mb genome reported by Keeling et al. (2013). The red palm weevil, *Rhynchophorus ferrugineus*, another pest from the curculionidae family, has a genome size of 782 Mb while the rice weevil, *Sitophilus oryzae*'s genome is 770 Mb (Hazzouri et al., 2020; Parisot et al., 2021). Assembled genome sizes of ground beetles are typically smaller compared to k-mer sequence-based methods, possibly because some repeated genomic regions remain unassembled (Pflug et al., 2020). In contrast, k-mer-based methods analyse the entire genome. A smaller assembly than expected may be due to incomplete or repeat collapses, while a larger assembly can be caused by independent haplotype assembly redundancies (Li et al., 2019). In brief, our *E. kamerunicus* genome size estimates using k-mer and flow cytometry align with coleopteran species' assembly sizes.

Obtaining precise estimations for complex genomes from actual sequencing data remains a challenging task to date. In future, efforts should be directed towards creating advanced algorithms that can effectively utilise genome sequences from other known insect species to enhance the accuracy of estimating genomic features. Fortunately, the *E. kamerunicus* genome is relatively small or falls within the intermediate size range among published insect genomes. This implies that future genome assembly and annotations for *E. kamerunicus* may be relatively simpler.

Genomic Features of Male and Female *E. kamerunicus* Genomes

K-mer analysis predicts genome size and reveals genomic features, like heterozygosity ratio and repeat rates. The GC content can be derived from the subsequent assembly analysis. Heterozygosity ratio, repeat rate and GC content are fundamental features for evaluating genome assembly quality and estimating genome size, particularly when using sequence-based approaches (Liu et al., 2013). Heterozygosity, which represents the proportion of differing nucleotides between an individual's inherited chromosomes from their parents, is a crucial measure for understanding genetic diversity (Bryc et al., 2013). Since the weevil samples used in this study are not inbred, the high heterozygosity level in the *E. kamerunicus* species (high heterozygosity ratio ≥0.8%, 0.5% ≤low heterozygosity <0.8%) is expected. High heterozygosity is common in various insect genomes, including Lepidoptera (Li et al., 2019). For example, the *Papilio glaucus* (Eastern tiger swallowtail) exhibits genome heterozygosity as high as 2% (Cong et al., 2015). The *Plutella xylostella* (diamond back moth) maintains a high level of heterozygosity even after ten generations of laboratory inbreeding. Remarkably, it was the first successfully sequenced insect genome that possess a significant degree of heterozygosity (You et al., 2013). However, our findings reveal that the male weevil has a 1.62% heterozygosity rate, while the female has a 1.92% rate (*Table 1*). Genome assembly may become challenging when genomic heterozygosity exceeds 1.00% (Zhou et al., 2022). Hence, the *E. kamerunicus* genome may be relatively difficult to assemble, as suggested by the short N50 scaffold lengths obtained (*Table 3*).

The genomes of both male and female *E. kamerunicus* belong to the category of low repetitive genomes, as their repeat ratios were below 50%. Specifically, the female *E. kamerunicus* has a repeat ratio of 43.49%, and the male has a repeat ratio of 43.70% (*Table 1*). The *E. kamerunicus* genome is considered a relatively simple repetitive genome, which may be linked to the species' small to medium genome size. High levels of heterozygosity, repeat sequences, and sequencing errors can reduce the precision of genome size estimation when utilising k-mer frequency (Liu et al., 2013). The high heterozygosity in the *E. kamerunicus* genome may explain the differences between our k-mer estimation and the cytometric approach. Genomic features statistics, such as repeat ratio, heterozygosity ratio, and GC content, were generally similar for both male and female *E. kamerunicus*.

High or low GC content exceeding the normal range (>65% or <25%) on the Illumina sequencing platform can introduce sequencing bias, potentially leading to reduced coverage of sequencing regions and affecting genome assembly (Zhou et al., 2022). Yet, GC content within the range of 30%-50% does not notably impact genome sequence quality (Shangguan et al., 2013). Both female and male *E. kamerunicus* genomes had similar GC content (*Figure 3*). The GC content of 31.71% in the male *E. kamerunicus* genome here was consistent with the male *E. kamerunicus* genome draft assembly (Apriyanto & Tambunan, 2021). Genomes of other curculionidae species also exhibit similar levels of GC content. *Dendroctonus ponderosae* and the model beetle *T. castaneum* had similar GC content (36% and 33%, respectively) (Keeling et al., 2013; Richards et al., 2008). The coffee borer beetle, Hypothenemus hampei, had a GC content of 32.46% (Vega et al., 2015). The red palm weevil, *R. ferrugineus*, had a GC content of about 32%, while the rice weevil *S. oryzae* had a GC content of 32.9% (Hazzouri et al., 2020; Parisot et al., 2021).

TABLE 3. GENOME ASSEMBLY STATISTICS FOR *E. kamerunicus*

Item	Sex	Total number of contigs and scaffolds	Maximum length (bp)	N50 length (bp)	N90 length (bp)
Contig	Female	2,837,454	79,424	310	117
	Male	2,631,773	85,498	280	110
Scaffold	Female	2,729,248	79,424	334	120
	Male	2,467,789	193,303	318	113

Note: bp - base pairs.

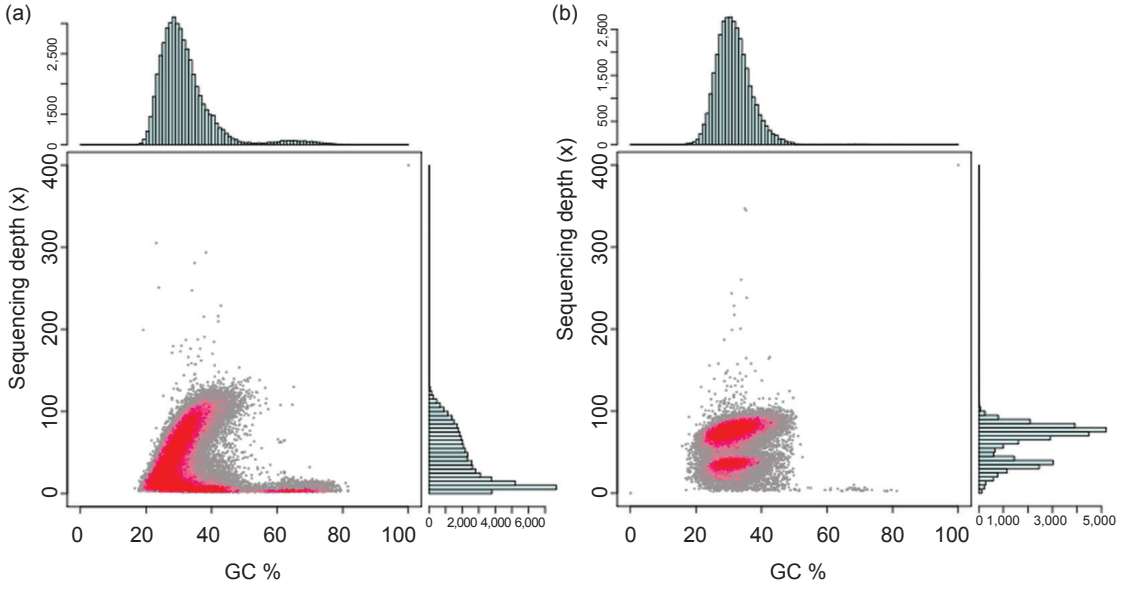

Figure 3. GC content and depth correlative analysis of (a) female, and (b) male E. kamerunicus. *Red regions represent relatively dense portions of the scatter plot. The right bar chart shows sequencing depth distribution, while the upper-left one represents GC content distribution.*

De novo Short-read Genome Assembly

A total of 213.84 and 215.16 Gb of raw sequence data were generated from female and male *E. kamerunicus*, respectively. Quality trimming and filtering, including a step to remove contaminating sequences, provided 63.90 Gb (female) and 64.50 Gb (male) of clean bases with Q20 scores >97%. The overall base error rate was a mere 0.03%. *Figure 4* shows the proportion of single bases, which is used to distinguish whether AT and GC separation is present. The results indicated good sequencing quality, with balanced A/G and C/T base content.

For assembly analysis, a k-mer of 41 was used due to the low repeat ratio of around 43% for both male and female *E. kamerunicus* (*Table 1*). From the female and male *E. kamerunicus*, 2.8 million and 2.6 million contigs were assembled, with N50 values of 310 and 280 bp, respectively. Subsequently, the draft female and male genome assemblies consist of 2.7 and 2.5 million scaffolds, with slightly higher N50 values of 334 and 318 bp, respectively (*Table 3*). The slight N50 value increase in the scaffold assemblies suggests no significant improvement over the contig assemblies.

The high heterozygosity rate (1.62%-1.92%) in the *E. kamerunicus* genome may affect genome size estimation and assembly quality. According to Li et al. (2020), the shorter N50 lengths of contigs and scaffolds (183 and 186 bp) in the Sichuan pepper genome may be due to its high heterozygosity rate of 1.73%. A BLASTN search was performed, examining 10,000 pairs of randomly processed clean reads from both male and female datasets for contamination assessment. BLAST results showed that *E. kamerunicus* sequences had their closest match in the mountain pine weevil *D. ponderosae*, with low homology (*Table 4*). These beetles are from different subfamilies. *Dendroctonus ponderosae* is in the Scolytinae subfamily, which comprises phloem-feeding beetles, while *E. kamerunicus* is in the curculioninae subfamily. Similar to *E. kamerunicus*, Acinetobacter bacterial sequences were found in the *D. ponderosae* assembly, indicating the presence of possible gut symbiont bacteria in coleopteran species (Keeling et al., 2013). However, Acinetobacter sequences in *E. kamerunicus* were relatively low, about 4%, and were only detected in the female *E. kamerunicus* sequencing data. The low percentages (3.2% to 4.5%) of *Wolbachia*, *Streptomyces*

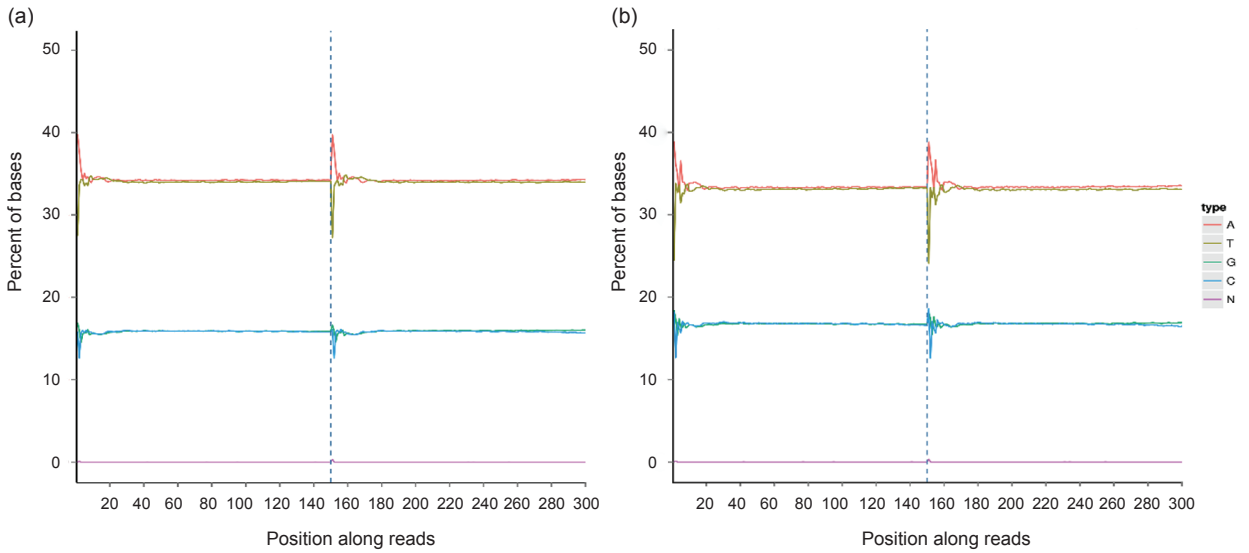

Figure 4. GC content distribution of (a) female, and (b) male E. kamerunicus.

TABLE 4. BLASTN HITS* STATISTICS FOR THE *E. kamerunicus* GENOME

Sex of *E. kamerunicus*	Hit 1	Hit 2	Hit 3	Hit 4	Hit 5
Female	*Dendroctonus ponderosae*, 20.79%	Wolbachia endosymbiont of *Culex quinquefasciatus Pel*, 4.49%	*Acinetobacter baumannii*, 3.93%	*Bombyx mori*, 3.37%	*Plautia stali* symbiont, 2.81%
Male	*Dendroctonus ponderosae*, 24.19%	*Homo sapiens*, 8.06%	*Bombyx mori*, 4.84%	*Streptomyces lividans* TK24, 4.03%	Wolbachia endosymbiont of *Culex quinquefasciatus Pel*, 3.23%

Note: *Hit: Species name and percentage of reads with hits annotated to it.

bacteria, *Bombyx mori* moth, and even *Homo sapiens* sequences in the *E. kamerunicus* data could be due to environmental contamination or minor sequence similarity among these species. These low contamination levels were deemed acceptable since they are consistent with findings in other insects, such as the coffee berry borer (Vega et al., 2015). Hence, *E. kamerunicus* genome data likely contains a small portion of cobiont genome sequences, possibly due to natural gut bacteria or cobionts within these weevils. These cobionts' genomes need to be considered when conducting genomic analysis such as de novo assembly. These factors, such as high heterozygosity, cobiont genomes, and pooled samples, coupled especially with the low coverage of the SkimSeq data, likely negatively affected the assembly results. While our assembly appears fragmented, future studies could benefit from combining long and short read data from various sequencing platforms, potentially improving and increasing the average sequence size used for assemblies (Amarasinghe et al., 2020).

CONCLUSION

Insects often show differences in genome size estimates when using sequence-based and flow cytometric methods. Not surprisingly therefore, genome size estimations for the male and female *E. kamerunicus* measured using these two approaches were different. The genome size estimate for the male *E. kamerunicus* is 202-381 Mb, with an almost similar range of 209-366 Mb for the female *E. kamerunicus*. The SkimSeq data also suggests *E. kamerunicus* has a simple repetitive genome with high heterozygosity rates. The current genome database for insects consists of only nine curculionids, thus our study contributes sequence data for the *E. kamerunicus* genome, encompassing both sexes.

ACKNOWLEDGEMENT

The authors would like to thank the Director-General of the Malaysian Palm Oil Board (MPOB) for permission to publish this article. We are grateful to the Universiti Kebangsaan Malaysia (UKM) for providing the *D. melanogaster* samples. We also thank Mr. Muhammad Azwan Zulkifli for the technical assistance on the BD FACSCalibur™ flowcytometry system (BD Biosciences). We acknowledge MPOB for the financial support of this study.

REFERENCES

Amarasinghe, S. L., Su, S., Dong, X., Zappia, L., Ritchie, M. E., & Gouil, Q. (2020). Opportunities and challenges in long-read sequencing data analysis. *Genome Biology*, *21*(1), 30. https://doi.org/10.1186/s13059-020-1935-5

Appiah, S. O., & Agyei Dwarko, D. (2013). Studies on entomophil pollination towards sustainable production and increased profitability in the oil palm: A review. *Elixir Agriculture*, *55*, 12878–12883.

Apriyanto, A., & Tambunan, V. B. (2021). Draft genome sequence, annotation, and SSR mining data of *Elaeidobius kamerunicus* Faust., an essential oil palm pollinating weevil. *Data in Brief*, *34*, 106745. https://doi.org/10.1016/j.dib.2021.106745

Barcelos, E., Rios, S. A., Cunha, R. N. V., Lopes, R., Motoike, S. Y., Babiychuk, E., Skirycz, A., & Kushnir, S. (2015). Oil palm natural diversity and the potential for yield improvement. *Frontiers in Plant Science*, *6*, 190. https://doi.org/10.3389/fpls.2015.00190

Bennett, M. D., Bhandol, P., & Leitch, I. J. (2000). Nuclear DNA amounts in angiosperms and their modern uses – 807 new estimates. *Annals of Botany*, *86*(4), 859–909. https://doi.org/10.1006/anbo.2000.1253

Biémont, C. (2008). Genome size evolution: Within-species variation in genome size. *Heredity*, *101*(4), 297–298. https://doi.org/10.1038/hdy.2008.80

Boulesteix, M., Weiss, M., & Biémont, C. (2006). Differences in genome size between closely related species: The *Drosophila melanogaster* species subgroup. *Molecular Biology and Evolution*, *23*(1), 162–167. https://doi.org/10.1093/molbev/msj012

Bryc, K., Patterson, N., & Reich, D. (2013). A novel approach to estimating heterozygosity from low-coverage genome sequence. *Genetics*, *195*(2), 553–561. https://doi.org/org/10.1534/genetics.113.154500

Bureš, P., Wang, Y. F., Horová, L., & Suda, J. (2004). Genome size variation in Central European species of *Cirsium* (Compositae) and their natural hybrids. *Annals of Botany*, *94*(3), 353–363. https://doi.org/10.1093/aob/mch151

Chen, W., Hasegawa, D. K., Arumuganathan, K., Simmons, A. M., Wintermantel, W. M., Fei, Z., & Ling, K. S. (2015). Estimation of the whitefly *Bemisia tabaci* genome size based on k-mer and flow cytometric analyses. *Insects*, *6*(3), 704–715. https://doi.org/10.3390/insects6030704

Cock, P. J. A., Fields, C. J., Goto, N., Heuer, M. L., & Rice, P. M. (2009). The Sanger FASTQ file format for sequences with quality scores, and the Solexa/Illumina FASTQ variants. *Nucleic Acids Research*, *38*(6), 1767–1771. https://doi.org/10.1093/nar/gkp1137

Cong, Q., Borek, D., Otwinowski, Z., & Grishin, N. V. (2015). Tiger swallowtail genome reveals mechanisms for speciation and caterpillar chemical defence. *Cell Reports*, *10*(6), 910–919. https://doi.org/10.1016/j.celrep.2015.01.026

Cunningham, C. B., Ji, L., Wiberg, R. A. W., Shelton, J., McKinney, E. C., Parker, D. J., Meagher, R. B., Benowitz, K. M., Roy-Zokan, E. M., Ritchie, M. G., Brown, S. J., Schmitz, R. J., & Moore, A. J. (2015). The genome and methylome of a beetle with complex social behaviour, *Nicrophorus vespilloides* (Coleoptera: Silphidae). *Genome Biology and Evolution*, *7*(12), 3383–3396. https://doi.org/10.1093/gbe/evv194

Dzulhelmi, M. N., Razi, A. N., Kalog, N. S. F., Murdi, A. F., Su, S., & Hazmi, I. R. (2022). Assessment on pollen carrying capacity and pollen viability of *Elaeidobius kamerunicus* (Coleoptera: Curculionidae). *Philippine Agricultural Scientist*, *105*(3), 309–314. https://doi.org/10.62550/abc003022

Galbraith, D. W., Harkins, K. R., Maddox, J. M., Ayres, N. M., Sharma, D. P., & Firoozabady, E. (1983). Rapid flow cytometric analysis of the cell cycle in intact plant tissues. *Science*, *220*, 1049–1051. https://doi.org/10.1126/science.220.4601.1049

Gregory, T. R. (2002). Genome size and developmental complexity. *Genetica*, *115*, 131–146. https://doi.org/10.1023/a:1016032400147

Gregory, T. R. (2023). *Animal genome size database*. Retrieved December 22, 2023, from http://www.genomesize.com

Guo, L. T., Wang, S. L., Wu, Q. J., Zhou, X. G., Xie, W., & Zhang, Y. J. (2015). Flow cytometry and k-mer analysis estimates of the genome sizes of *Bemisia tabaci* B and Q (Hemiptera: Aleyrodidae). *Frontiers in Physiology*, *6*, 144. https://doi.org/10.3389/fphys.2015.00144

Hanrahan, S. J., & Johnston, J. S. (2011). New genome size estimates of 134 species of arthropods.

Chromosome Research, *19*(6), 809–823. https://doi.org/10.1007/s10577-011-9231-6

Hazzouri, K. M., Sudalaimuthuasari, N., Kundu, B., Nelson, D., Al-Deeb, M. A., Le Mansour, A., Spencer, J. J., Desplan, C., & Amiri, K. M. A. (2020). The genome of pest *Rhynchophorus ferrugineus* reveals gene families important at the plant-beetle interface. *Communications Biology*, *3*, 323. https://doi.org/10.1038/s42003-020-1060-8

He, K., Lin, K., Wang, G., & Li, F. (2016). Genome sizes of nine insect species determined by flow cytometry and k-mer analysis. *Frontiers in Physiology*, *7*, 569. https://doi.org/10.3389/fphys.2016.00569

Herndon, N., Shelton, J., Gerischer, L., Ioannidis, P., Ninova, M., Dönitz, J., Waterhouse, R. M., Liang, C., Damm, C., Siemanowski, J., Kitzmann, P., Ulrich, J., Dippel, S., Oberhofer, G., Hu, Y., Schwirz, J., Schacht, M., Lehmann, S., Montino, A., . . . Bucher, G. (2020). Enhanced genome assembly and a new official gene set for *Tribolium castaneum*. *BMC Genomics*, *21*(1), 1–13. https://doi.org/10.1186/s12864-019-6394-6

Keeling, C. I., Yuen, M. M., Liao, N. Y., Docking, T. R., Chan, S. K., Taylor, G. A., Palmquist, D. L., Jackman, S. D., Nguyen, A., Li, M., Henderson, H., Janes, J. K., Zhao, Y., Pandoh, P., Moore, R., Sperling, F. A., Huber, D. P. W., Birol, I., Jones, S. J., & Bohlmann, J. (2013). Draft genome of the mountain pine beetle, *Dendroctonus ponderosae* Hopkins, a major forest pest. *Genome Biology*, *14*(3), R27. https://doi.org/10.1186/gb-2013-14-3-r27

Li, F., Zhao, X., Li, M., He, K., Huang, C., Zhou, Y., Li, Z., & Walters, J. R. (2019). Insect genomes: Progress and challenges. *Insect Molecular Biology*, *28*(6), 739–758. https://doi.org/10.1111/imb.12599

Li, J., Li, S., Kong, L., Wang, L., Wei, A., & Liu, Y. (2020). Genome survey of *Zanthoxylum bungeanum* and development of genomic-SSR markers in congeneric species. *Bioscience Reports*, *40*(6), BSR20201101. https://doi.org/10.1042/BSR20201101

Li, R., Zhu, H., Ruan, J., Qian, W., Fang, X., Shi, Z., Li, Y., Li, S., Shan, G., Kristiansen, K., Li, S., Yang, H., Wang, J., & Wang, J. (2010). *De novo* assembly of human genomes with massively parallel short read sequencing. *Genome Research*, *20*(2), 265–272. https://doi.org/10.1101/gr.097261.109

Liu, B., Shi, Y., Yuan, J., Hu, X., Zhang, H., Li, N., Li, Z., Chen, Y., Mu, D., & Fan, W. (2013). Estimation of genomic characteristics by analyzing k-mer frequency in *de novo* genome projects. *arXiv*, *1308*.2012. https://doi.org/10.48550/arXiv.1308.2012

Luo, R., Liu, B., Xie, Y., Li, Z., Huang, W., Yuan, J., He, G., Chen, Y., Pan, Q., Liu, Y., Tang, J., Wu, G., Zhang, H., Shi, Y., Liu, Y., Yu, C., Wang, B., Lu, Y., Han, C., . . . Wang, J. (2012). SOAPdenovo2: An empirically improved memory-efficient short-read *de novo* assembler. *GigaScience*, *1*(1), 18. https://doi.org/10.1186/s13742-015-0069-2

Marçais, G., & Kingsford, C. (2011). A fast, lock-free approach for efficient parallel counting of occurrences of k-mers. *Bioinformatics*, *27*(6), 764–770. https://doi.org/10.1093/bioinformatics/btr011

Meléndez, M. R., & Ponce, W. P. (2016). Pollination in the oil palms *Elaeis guineensis*, *E. oleifera* and their hybrids (OxG), in tropical America. *Pesquisa Agropecuária Tropical*, *46*(1), 102–110. https://doi.org/10.1590/1983-40632016v4638196

Meyer, J. M., Markov, G. V., Baskaran, P., Herrmann, M., Sommer, R. J., & Rödelsperger, C. (2016). Draft genome of the scarab beetle *Oryctes borbonicus* on la Réunion island. *Genome Biology and Evolution*, *8*(7), 2093–2105. https://doi.org/10.1093/gbe/evw133

Mohamad, S. A., Masri, M. M. M., Kamarudin, N., Sulaiman, M. R., Costa, A., Ong-Abdullah, M., Othman, H., Ahmad, S. N., Syarif, M. N. Y., Karim, Z. A., Abdul Ghani, I., Amit, S., Zakaria, A., Ming, S. C., Chuan, S. T., Jalinas, J., Koong, Y. Y., Sairi, A. A., Syed Ali, S. M. F., . . . Parveez, G. K. A. (2023). Impact of *Elaeidobius kamerunicus* (Faust) introduction on oil palm fruit formation in Malaysia and factors affecting its pollination efficiency: A review. *Journal of Oil Palm Research*, *35*(1), 1–22. https://doi.org/10.21894/jopr.2022.0021

Morgan-Richards, M., Trewick, S. A., Chapman, H. M., & Krahulcova, A. (2004). Interspecific hybridization among *Hieracium* species in New Zealand: Evidence from flow cytometry. *Heredity*, *93*(1), 34–42. https://doi.org/10.1038/sj.hdy.6800476

Mounsey, K. E., Willis, C., Burgess, S. T. G., Holt, D. C., McCarthy, J., & Fischer, K. (2012). Quantitative PCR-based genome size estimation of the astigmatid mites *Sarcoptes scabiei*, *Psoroptes ovis* and *Dermatophagoides pteronyssinus*. *Parasites & Vectors*, *5*, 3. https://doi.org/10.1186/1756-3305-5-3

Parisot, N., Vargas-Chávez, C., Goubert, C., Baa-Puyoulet, P., Balmand, S., Beranger, L., Blanc, C., Bonnamour, A., Boulesteix, M., Burlet, N., Calevro, F., Callaerts, P., Chancy, T., Charles, H., Colella, S., Da Silva Barbosa, A., Dell'Aglio, E., Di Genova, A., Febvay, G., . . . Heddi, A. (2021). The transposable element-rich genome of the cereal pest *Sitophilus oryzae*. *BMC Biology*, *19*(1), 241. https://doi.org/10.1186/s12915-021-01158-2

Parveez, G. K. A., Tarmizi, A. H. A., Sundram, S., Loh, S. K., Ong-Abdullah, M., Palam, K. D. P., Salleh, K. M., Ishak, S. M., & Idris, Z. (2021). Oil palm economic performance in Malaysia and R&D progress in 2020. *Journal of Oil Palm Research*, *33*(2), 181–214. https://doi.org/10.21894/jopr.2021.0026

Parveez, G. K. A., Kamil, N. N., Zawawi, N. Z. I. N., Ong-Abdullah, M., Rasuddin, R., Loh, S. K., Selvaduray, K. R., Hoong, S. S. O. I., & Idris, Z. (2022). Oil palm economic performance in Malaysia and R&D progress in 2021. *Journal of Oil Palm Research*, *34*(2), 185–218. https://doi.org/10.21894/jopr.2022.0036

Pflug, J. M., Holmes, V. R., Burrus, C., Johnston, J. S., & Maddison, D. R. (2020). Measuring genome sizes using read-depth, k-mers, and flow cytometry: Methodological comparisons in beetles (Coleoptera). *G3: Genes | Genomes | Genetics*, *10*(9), 3047–3060. https://doi.org/10.1534/g3.120.401028

Powell, D., Groβe-Wilde, E., Krokene, P., Roy, A., Chakraborty, A., Löfstedt, C., Vogel, H., Andersson, M. N., & Schlyter, F. (2021). A highly-contiguous genome assembly of the Eurasian spruce bark beetle, *Ips typographus*, provides insight into a major forest pest. *Communications Biology*, *4*(1), 1–9. https://doi.org/10.1038/s42003-021-02602-3

Richards, S., Gibbs, R. A., Weinstock, G. M., Brown, S. J., Denell, R., Beeman, R. W., Gibbs, R., Bucher, G., Friedrich, M., Grimmelikhuijzen, C. J. P., Klingler, M., Lorenzen, M., Roth, S., Schröder, R., Tautz, D., Zdobnov, E. M., Muzny, D., Attaway, T., Bell, S., . . . Kapustin, Y. (2008). The genome of the model beetle and pest *Tribolium castaneum*. *Nature*, *452*(7190), 949–955. https://doi.org/10.1038/nature06784

Sarmashghi, S., Bohmann, K., Gilbert, P. M. T., Bafna, V., & Mirarab, S. (2019). Skmer: Assembly-free and alignment-free sample identification using genome skims. *Genome Biology*, *20*(1), 34. https://doi.org/10.1186/s13059-019-1632-4

Shangguan, L., Han, J., Kayesh, E., Sun, X., Zhang, C., Pervaiz, T., Wen, X., & Fang, J. (2013). Evaluation of genome sequencing quality in selected plant species using expressed sequence tags. *PLoS ONE*, *8*(7), e69890. https://doi.org/10.1371/journal.pone.0069890

Vega, F. E., Brown, S. M., Chen, H., Shen, E., Nair, M. B., Ceja-Navarro, J. A., Brodie, E. L., Infante, F., Dowd, P. F., & Pain, A. (2015). Draft genome of the most devastating insect pest of coffee worldwide: The coffee berry borer, *Hypothenemus hampei*. *Scientific Reports*, *5*, 12525. https://doi.org/10.1038/srep12525

Woittiez, L. S., van Wijk, M. T., Slingerland, M., van Noordwijk, M., & Giller, K. E. (2017). Yield gaps in oil palm: A quantitative review of contributing factors. *European Journal of Agronomy*, *83*, 57–77. https://doi.org/10.1016/j.eja.2016.11.002

You, M., Yue, Z., He, W., Yang, X., Yang, G., Xie, M., Zhan, D., Baxter, S. W., Vasseur, L., Gurr, G. M., Douglas, C. J., Bai, J., Wang, P., Cui, K., Huang, S., Li, X., Zhou, Q., Wu, Z., Chen, Q., . . . Wang, J. (2013). A heterozygous moth genome provides insights into herbivory and detoxification. *Nature Genetics*, *45*(2), 220–225. https://doi.org/10.1038/ng.2524

Yu, Y. S., Jin, S., Cho, N., Lim, J., Kim, C. H., Lee, S. G., Kim, S., Park, J. S., Kim, K., Park, C., & Cho, S. J. (2021). Genome size estimation of *Callipogon relictus* Semenov (Coleoptera: Cerambycidae), an endangered species and a Korea natural monument. *Insects*, *12*(2), 111. https://doi.org/10.3390/insects12020111

Zhou, P., Li, J., Huang, J., Li, F., Zhang, Q., & Zhang, M. (2022). Genome survey sequencing and genetic background characterization of *Ilex chinensis* Sims (Aquifoliaceae) based on next-generation sequencing. *Plants*, *11*(23), 3322. https://doi.org/10.3390/plants11233322

Zonneveld, B. J. M. (2001). Nuclear DNA contents of all species of *Helleborus* (Ranunculaceae) discriminate between species and sectional divisions. *Plant Systematics and Evolution*, *229*(1), 125–130. https://doi.org/10.1007/s006060170022