Genome Survey of the Japanese pine sawyer beetle, *Monochamus alternatus* (Coleoptera: Cerambycidae)

Abstract: *Monochamus alternatus* is a longicorn beetle, an important stem borer of pine trees in China, and a major vector of pine wood nematode disease. In this study, we determined the first accurate genomic survey of *M. alternatus* by the next-generation sequencing (NGS), which is a relatively new methodology that can ensure the identification of large numbers of simple sequences repeat (SSR) markers. Our result showed that the genome size was 871.09 Mb, the GC content of the *M. alternatus* genome was within normal limits (34.45%), the proportion of repetitive sequences was high (59.40%), and the heterozygosity rate was low (1.04%). Dinucleotide repeats (65.55%) were the dominant form of SSR, and virtually all preponderant types in each nucleotide repeat consisted of Adenine (A) and Thymine (T).

Keywords: *Monochamus alternatus*, Coleoptera, genome survey, high-throughput sequencing.

Introdcution

Monochamus alternatus, belonging to the subfamily Lamiinae, is not only an important stem borer of pine trees in China, but also the main vector of pine wood nematode disease ((Nie et al. 2020)). *M. alternatus* plays a key role in carrying, spreading and assisting the pathogen to invade the host in the process of spread and infection of pine wood nematode (Wang et al. 2019). Natural enemies include pathogenic microorganisms, parasitic nematodes, parasitic insects, predatory insects, spiders, birds and so on. The life cycle can be divided into four stages: pupa, larva, egg and adult, with one generation a year, and the mature larva overwinters in the tunnel. Distributed in China, South Korea, Japan, Laos. It was listed on the list of forestry dangerous pests of the State Forestry Administration of China in 2003 (Santana et al. 2020). In terms of its genetic characteristics, the whole genome of *M. alternatus* has not yet been published (Wen et al. 2018, Ming et al. 2014). In view of the great harm of this species to forestry, we use the NGS technology to determine and analyze its whole genome in order to provide basic molecular data for further study of its behavior, physiology, ecology, genetic variation, evolutionary adaptation , population protection and the formulation of conservation strategies (Chen et al. 2012, Haddad et al. 2018, Kimberly et al. 2020, Dai et al. 2020, Kim et al. 2020, Hu et al. 2013, Yang et al. 2003, Xu et al. 2013,).

Despite its great harmless to pine wood, the genetic information of *M. alternatus* has remained largely unknown, the genome size and genome-wide sequencing of *M. alternatus* have not been reported, the mitogenomes of *M. alternatus Bursaphelenchus mucronatus*, and some of the closely related species have been studied (Shi et al. 2015). Formerly, scholars have used morphology, anatomy and physiology, and other methods to establish relationships among species for creating classification systems. However, the evolutionary relationships between *M. alternatus* and other

closely related species are hard to describe just based on traditional dates. Previous studies have used mitogenome dates to assess the gene expression of *M. alternatus* (Bai et al. 2015), thus, we aim to expand the database of Lamiinae and provide some useful information for further study.

Genome is the general name of all genetic information of a particular organism, which ultimately determines the transmission of genetic material, while genomics can be defined as a science that studies the structure, function and diversity of genomes. With the development of the sequencing techniques, more and more manuscript genomes have been assembled. The main difference between genomics and traditional biological methods lies in the scale of research, because the goal of genomics is to analyze a large number of genes extensively and may even involve a complete set of genes that make up the genome, rather than being limited to one or a small number of genes. With the development of DNA sequence technology, the cost has been decreased rapidly and the ability of obtaining the date has been increased vastly.

Next-generation sequencing (NGS) technology has significantly increased the sequence output while reducing time and cost. Genomic survey is a method that combine NGS technology with *K*-mer dates to achieve species genome size, GC content, heterozygosity rate and repetition rate. This technology has been used widely and could also accurately predict the whole genome sizes.

In this study, we provided valuable molecular data to reveal the relationship between *M. alternatus* and pine wood nematode, Zhou et al. (2018) proved that *M. alternatus* infected with *B. xylophilus* showed an increase in the expression of some antioxidant genes, so that it could obtain immune tolerance, this study showed that the existence of *B. xylophilus* would increase the expression level of some metabolic (Timothy et al. 2018) , so some scholars inferred that the relationship between nematodes and beetles should be host switching rather than nematode-beetle

coevolution, because in their study, they found some efforts were made by *M. alternatus* to offset the impact. The mechanisms need to be studied in the future.

Methods and Materials

During the inspection in Anxi county of Quanzhou, we had obtained the *M. alternatus* specimens, which were frozen in liquid nitrogen and stored in a cryogenic refrigerator at -80 □.

Sample extraction and detection

The genomic DNA, was extracted by improved SDS method. The OD values of 260 nm and 280 nm were determined by ultraviolet spectrophotometer to estimate the concentration and purity of the extracted DNA, and then agarose gel electrophoresis was performed to detect the integrity of the extracted DNA. When sampling, 1 kb DNA Ladder (Takara) and λ-Hind □ di-gest (Takara) were used as Marker, gel concentration of 0.8%, electrophoresis for 60 minutes under the condition of 80V voltage. After the electrophoresis was completed, the electrophoresis of genomic DNA was observed in the gel imaging system of BIO-RAD company.

Sequencing

The DNA samples were randomly interrupted by Covaris ultrasonic crusher, and the small fragment sequencing library of 250bp was established. Then the whole library was prepared by terminal repair, A tail, sequencing joint, purification, PCR amplification and so on. After that, double-terminal (Pair-End) sequencing was carried out on the Illumina HiseqTM2000 platform. Finally, the quality control is carried out by using the software NGS-QC-Generator, the low-quality data is filtered out, and the effective data obtained are used for genome feature evaluation and

preliminary assembly.

17-mer Analysis

K-mer analysis was used to estimate the genome size. Use the formula: genome size = *K*-mer number / peak depth to estimate genome size. The estimated value of *K*-mer depth is calculated, which is used to estimate the genome size. The heterozygosity (Φ) was calculated by the following formula, while the repetition rate was calculated by calculating the ratio of the number of *K*-mer to the total number after 1.8 times the depth of the homozygous peak. In the formula, *K* is the number of heterozygous *K*-mer, A1 *K*-mer 2 is the percentage of heterozygous *K*-mer species, and n*K* species is the number of all *K*-mer species.

Distribution and Analysis of GC content

The GC depth content of the target genome was calculated by constructing the GC dot map from the total number of bases in the sequencing data, and the correlation analysis was carried out. Contigs were rearranged by all clean reads and scaffolds were gradually constructed by diversified insert size paired-ends.

Preliminary Assembly of Genome

The sequence reads obtained from all the small fragment libraries was truncated into smaller sequence fragments, and the de-Brujin map was constructed by using the overlapping relationship between them. Then the simplified de-Brujin map which is easy to analyze was obtained by screening branches, simplifying the bifurcation pathway and randomly combining heterozygous

sites, and then the bifurcation sequence was truncated into small fragments to get the initial overlapping (contigs). The reads obtained by sequencing was compared, and the obtained contigs, assembled the contigs into scaffolds by using the connection relationship between reads and the size information of insertion fragments. In order to make the assembled sequence more complete, it is necessary to connect the contigs, according to the pairing relationship between the double-terminal data and optimize and fill holes in the gaps between the contig, so as to obtain the original genome sequence.

Results

Sequencing Quality

The low-quality data is filtered out by screening, and the effective data with the size of 34.68Gb is obtained, and the sequencing depth is 51x. Through the strict filtering of the obtained data, the high-quality pure data is obtained. The 350 bp library is constructed, the original data is 34,683,801,900bp, and the clean data 34,652,586,478bp is obtained after filtering. Both Q20 and Q30 are indicators to measure the quality of sequencing. It is generally believed that when Q20 ≥ 90% and Q30 ≥ 80%, the quality of sequencing data is better. In this study, the content of Q20 is 97.86% and the content of Q30 is 93.48%. It is known that the sequencing quality of this study is good, and the sequencing error rate is 0.04%, which is also within the normal range (< 0.05%).

After genome regulation, if the estimated heterozygosity of the genome is high (higher than 0.5%), or the content of repetitive sequences is high (more than 50%), the genome can be regarded as a complex genome. WGS sequencing and assembly strategies suitable for general genomes are generally difficult to obtain good complex genome maps. At present, there are two commonly used

strategies to solve complex genome maps, SangeH-454 sequencing and WGS+BAC to BAC/fosmid to fosmid. Compared with the former, the latter is based on Illumina sequencing, and the cost is much lower. In this study, we found the heterozygosity of the *M. alternatus* is higher than 0.5%, the repeat rate is more than 50% either. In consequence, the use of Illumina+PacBio in sequencing assembly strategy is highly recommended.

17-mer Analysis and Qenome Size Estimation

Genome size of *M. alternatus* was estimated according to the Lander waterman algorithm based on *K*-mer (k = 17) frequency of the clean reads and the 17-mer frequency distribution complied with the Poisson distribution. The 34.65Gb valid data of *M. alternatus* were analyzed by *K*-mer (figure 1). It can be seen that the peak is near depth=29; the calculated genome size is 871.09 Mbp, the modified genome size is 860.06 Mbp. There is a tail on the *K*-mer curve, indicating that the proportion of genome repeat sequences of *M. alternatus* is large.

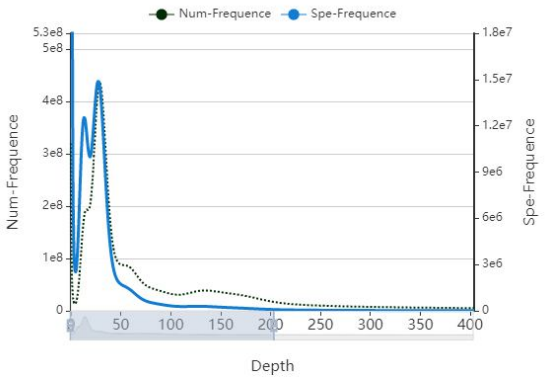

Fig. 1. 17 *K*-mer analysis for estimating the genome size of *Monochamus alternatus*
Note: The genome size was estimated by using the formula: Genome size=*K*-me num/Peak depth

Results of Preliminary Assembly of the Genome

Using SOAP-Denovo software to assemble the data, the initial genome sequence is obtained, and

the splicing result is shown in Table 1. It can be seen that the genomic Contig N50 of *M. alternatus* is 787bp and scaffolding N50 is 954bp, which is short, so it is not suitable for further assembly by shotgun method.

Table 1. Statistics of assembly results in the *Monochamus alternatus* genome

Item	Conting	Scaffold
	Size (bp)	Size (bp)
N50	787	954
N90	135	145
Total Size	823 399 058	840 047 447
Total Number	2 034 890	1 839 509
Longest	385 511	458 244

Content and Distribution of GC

GC content is an important character of nucleic acid sequence composition in living creatures, which is also used to detect the separation of AT and GC. Theoretically, the contents of G and C bases and An and T bases should be equal in each sequencing cycle, and the whole sequencing process is stable and horizontal. In the actual sequencing process, due to the DNA template amplification deviation and the low sequencing quality value of the first few bases, it often leads to large fluctuations in the first few bases of each read, which is a normal situation. According to the statistics of GC content of assembled contig, the distribution map of GC content and sequencing depth (depth) was obtained (Fig. 2). The genomic GC content of *M. alternatus* was 32.94%.

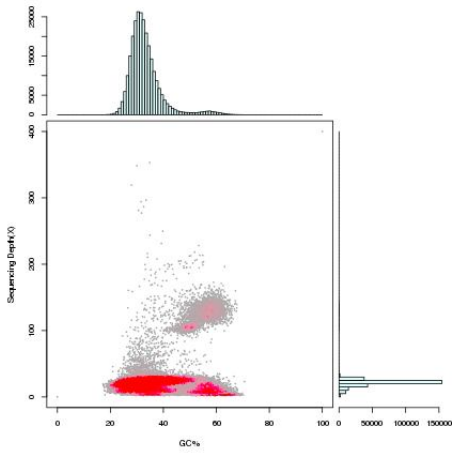

Fig. 2. GC content and average sequencing depth

Genome Comparison of the Insects in Coleoptera

In the genome database of the National Biotechnology Information Center of the United States (https://www.ncbi.nlm.nih.gov/genome), enter the key words "Coleoptera" to obtain 61 pieces of genome information of Coleoptera insects. We select 25 species that can be reported in the relevant literature, and compare their genome information with the genome data of *M. alternatus* (Table 2). According to the published data, the genome size of Cerambycidae ranges from 90.87 to 1112.44 Mb, while the genome size of *M. alternatus* is as high as 871.09 Mb. Some studies believe that if the content of genomic GC is moderate (25% < GC < 65%), it will not affect the accuracy of genome sequencing and correct assembly (Airdetal. 2011). The content of genomic GC measured in this study is 32.94%, so the sequencing results and assembly should be correct. On the other hand, the GC content of the other 24 genomes was similar.

Table 2. Comparison of the mitogenome assembly data of *Monochamus alternatus* with that of 25 species of Coleoptera

Species	Genome size (Mb)	Number of protein-coding genes	GC content (%)	Heterozygosity (%)	Repeat (%)	References
Monochamus alternatus	871.09	n.a.	32.94	1.04	59.04	This study
Diabrotica virgifera	2418.07	28061	36.5	n.a.	n.a.	Wu K,et al.,2017
Photinus pyralis	471.51	32294	36.4	0.598	42.6	Fallon TR et al.,2018
Listronotus bonariensis	1112.44	n.a.	31.3	0.18	70	Harrop TWR, et al.,2020
Propylea japonica	851.23	18018	35.13	0.9	58.22	Zhang L, et al.,2020
Marronus borbonicus	406.94	23278	35.9	0.2	29.2	Meyer JM, et al.,2016
Abscondita terminalis	499.65	20439	31.4	n.a.	n.a.	Chen X, et al.,2019
Lamprigera yunnana	1052.93	19438	34.1	n.a.	n.a.	Chen X, et al.,2019
Hycleus cichorii	99.17	13813	32.3	1.00	72.73	Wu YM, et al.，2018
Harmonia axyridis	466.692	n.a.	33.2	n.a.	n.a.	Ando T, et al.，2018
Hycleus phaleratus	90.87	13725	30.06	0.99	74.90	Du C, et al.,2017
Aethina tumida	234.34	17463	30.00	n.a.	n.a.	Evans JD, et al.,2018
Limonius californicus	1072.67	n.a.	35.00	0.21	n.a.	Andrews KR, et al.,2020
Oryctes borbonicus	406.20	8822	34.85	0.2	29.2	Meyer JM, et al.,2016
Nicrophorus vespilloides	195.27	18995	32.20	n.a.	n.a.	Cunningham CB, et al.,2015
Asbolus verrucosus	249.61	n.a.	32.90	n.a.	n.a.	Stanley Dean Rider Jr,2015

Protaetia brevitarsis	1143.92	n.a.	25.40	n.a.	72.29	Kim MJ, et al.,2014
Popillia japonica	531.53	n.a.	34.90	n.a.	n.a.	Yang W, et al.,2018
Anoplophora glabripennis	706.97	20632	33.40	n.a.	n.a.	Lowe TM, et al.,1997
Tenebrio molitor	280.78	n.a.	26.50	n.a.	n.a.	Liu LN, et al.,2014
Agrilus planipennis	353.07	22159	36.00	n.a.	n.a.	Lord NP, et al.,2016
Leptinotarsa decemlineata	641.99	19038	35.60	n.a.	n.a.	Sharkey CR, et al.,2017
Callosobruchus maculatus	1007.82	31345	37.7	n.a.	65.00	Sayadi A, et al.,2019
Priacma serrata	12.08	n.a.	35.90	n.a.	n.a.	Niehuis O, et al.,2012
Dendroctonus ponderosae	257.09	16791	38.45	n.a.	n.a.	Keeling CI, et al.,2016

Discussion

Although the depth of our *M. alternatus* preliminary sequencing genome is relatively low and it's a disadvantage for an acquaintance of genome information, we still excavated some basic genome information and developed genome resources by using several bioinformation methods like *K*-mer and scaffold assembly methods. According to each analysis index, it is inferred that the genome of *M. alternatus* is a highly complex genome, and the corresponding strategy can be used to assemble the genome.

There are many species of insects, about 1.8 million species are known, accounting for more than 3 beat 4 of the described animal species in the world (Wu et al. 2017, Sharkey et al. 2016, Lord et al. 2016, Nakamatsu et al. 2004). In recent years, with the continuous improvement of DNA sequencing technology in the direction of automation (Shah et al. 2020), generalization and low cost,

the continuous maturity of genomics research technology and the continuous updating of bioinformatics analysis methods (Wu et al. 2020), many entomologists use various ensemble research methods such as genome, transcriptome, proteome, etc. to obtain a large number of molecular biological data, and use bioinformatics methods to analyze and mine the data. This paper attempts to reveal the genetic basis of the special physiological activities and behaviors of some insects from the point of view of molecular biology (Bonasio et al. 2010, Cunningham et al. 2015), expounding the molecular mechanism of co-evolution of insects and other organisms (Wang et al. 2014, Oliver et al. 2012), showing a broad research prospect of insect molecular biology (Ren et al. 2020, Todd et al. 1997).

Coleoptera is the first order with the largest number of species and the widest distribution in the Insecta and even the animal kingdom. There are many kinds and the system is complex. However, there are also many species (mainly larvae) that harm crops and bring loss or inconvenience to human production and life (Zhang et al. 2019). Therefore, the research on Coleoptera insect genomics has important theoretical significance and application value. Nowadays, the genomic research of Coleoptera insects is mainly to analyze their special habits or behaviors at the genome level (Zhan et al. 2011, You et al. 2013, Zhan et al. 2014, Harrop et al. 2020). Genome size, also known as C-value or Constant-value, is an important genetic feature of an organism. *M. alternatus* is known as the cancer of pine trees and is also a devastating disease, causing serious environmental and economic losses. Effective strategies are needed to stop or control the spreading of this disease. This study will provide a method to assemble the genome and may serve as key point to develop new control strategies for pine wilt disease, the results of the current study will play an important role in future whole-genome sequencing projects and provide an abundant resource for further

functional studies (Jan et al. 2016, Yuan et al. 2018, Chao et al. 2017), which will help us learn genetic regulatory mechanisms.

This study lays the foundation of whole-genome sequencing in *M. alternatus*, and puts forward the platform for the next investigation on this particular insect. Today, more and more scholars studied the relationship between *M. alternatus* and pine wood nematode, because of its great harm to the pine wood. Our study increases the abundance of the Cerambycidae genome information and can assist in phylogenetic, molecular systematics and evolutionary studies of Cerambycidae. However, molecular data on species of Lamiinae are still scarce and more information is needed to fully explore the relationship within Lamiinae.

References

Nie, R., A.P. Vogler, X.K. Yang, and M.Y. Lin. 2020. Higher-level phylogeny of longhorn beetles (Coleoptera: Chrysomeloidea) inferred from mitochondrial genomes. Syst. Entomol. 46: 56–70.

Wang, J., X.Y. Dai, X.D. Xu, Z.Y. Zhang, D.N. Yu, K.B. Storey, and J.Y. Zhang. 2019. The complete mitochondrial genomes of five longicorn beetles (Coleoptera: Cerambycidae) and phylogenetic relationships within Cerambycidae. PeerJ. 7: e7633.

Santana, D.S., L. Marinoni, M.L. Monné, and J. Gómez-Zurita. 2020. Molecular phylogenetic assessment of the tribal classification of Lamiinae (Coleoptera: Cerambycidae). Mol. Phylogenet. Evol. 145: 106736.

Wen, Z., Y. Z. Yang, Q. F. Shi, J. L. Li, and H. L. Zhi. 2018. The first complete mitochondrial genome of the Japanese beetle *Popillia japonica* (Coleoptera: Scarabaeidae) and its phylogenetic implications for the superfamily Scarabaeoidea. Int. J. Biol. Macromol. 118:1406–1413.

Min, J., H. H. I. Kim, Y. L. Kwang, S. H. Yeon, and K. Iksoo. 2014. Complete mitochondrial genome of the whiter-spotted flower chafer, *Protaetia brevitarsis* (Coleoptera: Scarabaeidae). Mitochondrial DNA 25(3): 177–178.

Chen, W. B., D. N. Xiao, and X. Z. Li. 2012. Classification, application, and creation of landscape indices J. Appl. Entomol. 13(01) : 121-125.

Haddad, S., S. Shin, A.R. Lemmon, E.M. Lemmon, P. Svacha, B. Farrell, D. Windsor, and D.D. Mckenna. 2018. Anchored hybrid enrichment provides new insights into the phylogeny and evolution of longhorned beetles (Cerambycidae). Syst. Entomol. 43: 68–89.

Andrews, K.R., A. Gerritsen, A. Rashed, D.W. Crowder, S.I. Rondon, W.G. van Herk, R. Vernon, K.W. Wanner, C.M. Wilson, D.D. New, *et al.* **2020**. Wireworm (Coleoptera: Elateridae) Genomic analysis reveals putative cryptic species, population structure, and adaptation to pest control. Commun. Biol. 3: 489.

Dai, X.Y., H. Zhang, X.D. Xu, Y.Y. Jia, J.Y. Zhang, D.N. Yu, and H.Y. Cheng. 2020. The complete mitochondrial genome of *Annamanum lunulatum* (Coleoptera: Lamiinae) and its phylogeny. Mitochondrial DNA Part B 5: 551–553.

Zhang, Z.Y., J.Y. Guan, Y.R. Cao, X.Y. Dai, K.B. Storey, D.N. Yu, and J.Y. Zhang,. Mitogenome analysis of four Lamiinae Species (Coleoptera: Cerambycidae) and gene expression responses by *Monochamus alternatus* when infected with the parasitic nematode. Insects 12(05):453.

Hu, S. J., T. Ning，D.Y. Fu, R.A. Haack, Z. Zhang, D.D. Chen, X.Y. Ma, and H. Ye.. 2013. Dispersal of the Japanese pine sawyer, *Monochamus alternatus* (Coleoptera: Cerambycidae), in mainland China as inferred from molecular data and associations to indices of human activity. PLoS ONE 8(02): e57568.

Yang, F., and D. H. He 2003. Effects of habitat fragmentation on biodiversity. Ann. Rev. Ecol. Evol. S. 34(02): 487-515.

Xu, X. M. 2013. Effects of environmental conditions on infection of Fusarium ear blight. Eur. J. Plant Pathol. 109(07): 683-389.

Shi, H., C. Q. Zhang, L. Shan, and Q. Qi. 2015. First report of *Pestalo-tiopsis microspora* as a causal agent of black spot of pecan (*Carya illinoensis*) in China. Plant Dis. 9(09): 1276.

Bai, L., C. M. Tian，C. H. Hong，F.F. Kang, J.Y. Chen, D.W. Song, and H.G. Liu. 2015. The relationship between pine forest landscape patterns and pine wilt disease in Yichang of Hubei Province. Acta Ecologica Sinica. 35(24) : 8107-8116.

Fallon, T.R., S.E. Lower, C.H. Chang, M. Bessho-Uehara, G.J. Martin, A.J. Bewick, M. Behringer, H.J. Debat, I. Wong, J.C. Day, *et al.* **2018**. Firefly genomes illuminate parallel origins of bioluminescence in beetles. eLIFE. **7**: 4-5.

Jan, M., G. V. Meyer, P. B. Markov, H. Matthias, R. J. Sommer, and R. Christian. 2016. Draft genome of the scarab *Beetle Oryctes* borbonicus on La Re´union Island. Genome Biol. Evol. 8(07):2093–2105.

Yuan, M. W., L. Jiang, and S. C. Xiang. 2018. Draft genomes of two blister beetles *Hycleus cichorii* and *Hycleus phaleratus*. GigaScience 7: 1–7.

Chao, D., F. Z. Li, L. Ting, , N. M. Jing, J. Z. Chen, S. Y. Bi, and Y. Z. Xiu. 2017. Mitochondrial genomes of *Blister beetles* (Coleoptera, Meloidae) and two large intergenic spacers in *Hycleus* genera. BMC. Genomics. 18: 698.

Cunningham, C. B., L. Ji, R. A. Wiberg, J. Shelton, E.C. McKinney, D.J. Parker, R.B. Meagher, K.M. Benowitz, E.M. Roy-Zokan, M.G. Ritchie, *et al.* **2015**. The Genome and Methylome of a Beetle with Complex Social Behavior, *Nicrophorus vespilloides* (Coleoptera: Silphidae). Genome Biol. Evol. **7**(12): 3383–3396.

Todd, M., Lowe, and R. E. Sean. 1997. tRNAscan-SE: a program for improved detection of transfer RNA genes in genomic sequence. Nucleic Acids Res. 25(05): 955–964.

Niehuis O., G. Hartig, S. Grath, H. Pohl, J. Lehmann, H. Tafer, A. Donath, V. Krauss, C. Eisenhardt, J. Hertel, *et al*. **2012**. Genomic and Morphological Evidence Converge to Resolve the Enigma of Strepsiptera. Curr. Biol. 22: 1309–1313.

Wu, K., C. Camargo, E. Fishilevich, , K.E. Narva, X. Chen, C.E. Taylor, and B.D. Siegfried. 2017. Distinct fitness costs associated with the knockdown of RNAi pathway genes in western *Corn rootworm* adults. PLoS ONE 12(12): e0190208.

Sharkey, C.R., M.S. Fujimoto, N.P. Lord, S. Shin, D.D. McKenna, A. Suvorov, G.J. Martin, and S.M. Bybee. 2016. Overcoming the loss of blue sensitivity through opsin duplication in the largest animal group, beetles. Sci. Rep. **7**: 8.

Lord, N.P., R.L. Plimpton, C.R. Sharkey, A. Suvorov, J.P. Lelito, B.M. Willardson, and S.M. Bybee. 2016. A cure for the blues: opsin duplication and subfunctionalization for short-wavelength sensitivity in jewel beetles (Coleoptera: Buprestidae). BMC Evol. Biol. 16: 107.

Nakamatsu, Y. and T. Tanaka. 2004. The function of a trypsin—like enzyme in the saliva of *Euplectrus separatae* larvae. J. Insect Physiol. 50(09): 847-854.

Shah, K., M.A. Alqudah, F. Jarad, and T. Abdeljawad. 2020. Semi-analytical study of Pine Wilt Disease model with convex rate under Caputo–Febrizio fractional order derivative. Chaos. Soliton. Fract. 135: 109754.

Wu, S., S. Gao, S. Wang, J. Meng, J. Wickham, S. Luo, X. Tan, H. Yu, Y. Xiang, S. Hu, *et al*.2020. A reference genome of *Bursaphelenchus mucronatus* provides new resources for revealing its displacement by pinewood nematode. Genes. 11: 570.

Bonasio, R., G. Zhang, C. Ye, N.S. Mutti, X. Fang, N. Qin, G. Donahue, P. Yang, Q. Li, C. Li, *et al*. **2010**. Genomic comparison of the ants *Camponotus floridanus* and *Harpegnathos saltator*. Science 329(5995): 1068-1071.

Wang, X., X. Fang, P. Yang, X. Jiang, F. Jiang, D. Zhao, B. Li, F. Cui, J. Wei, C. Ma, *et al*. **2014**. The locust genome provides insight into swarm formation and long-distance flight. Nat. Commun. 5: 2957-2965.

Ren, B. Z., and L. H. Tang. 2020. Progress in quarantine pests of *Bursaphelenchus xylophilus*. J. Jilin Agric. Univ. 42: 8–13.

Zhang, J.Y., B.E. Luu, D.N. Yu, L.P. Zhang, R. Al-attar, and K.B. Storey. 2019. The complete mitochondrial genome of *Dryophytes versicolor*: Phylogenetic relationship among Hylidae and mitochondrial protein-coding gene expression in response to freezing and anoxia. Int. J. Biol. Macromol. 132: 461–469.

Zhan, S., C. Merlin, J.L. Boore, and S.M. Reppert. 2011. The monarch butterfly genome yields insights into long-distance migration. Cell 147(05): 1171-1185

You, M.S., Z. Yue, W. He, X. Yang, G. Yang, M. Xie, D. Zhan, S.W. Baxter, L. Vasseur, G.M. Gurr, *et al*. **2013**. A heterozygous moth genome provides insights into herbivory and detoxification. Nat. Genet. 45(02): 220-225.

Zhan, S., W. Zhang, K. Niitepõld, J. Hsu, J.F. Haeger, M.P. Zalucki, S. Altizer, J.C. de Roode, S.M. Reppert, and M.R. Kronforst. 2014. The genetics of monarch butterfly migration and warning colouration. Nature 514(7522): 317-31

Harrop, T.W.R., M.F. Le Lec, R. Jauregui, S.E. Taylor, S.N. Inwood, T. van Stijn, H. Henry, J. Skelly, S**. Ganesh, R.L. Ashby,** *et al*. **2020**. Genetic diversity in invasive populations of *Argentine Stem Weevil* associated with adaptation to biocontrol. Insects 11: 441.

Zhou, J., L. L. Zhao, H.Y. Yu, Y.H. Wang, W. Zhang, S.N. Hu, Z. Zou, and J.H. Sun. 2018. Immune tolerance of vector beetle to its partner plant parasitic nematode modulated by its insect parasitic nematode. FASEB J. 32: 4862–4877.

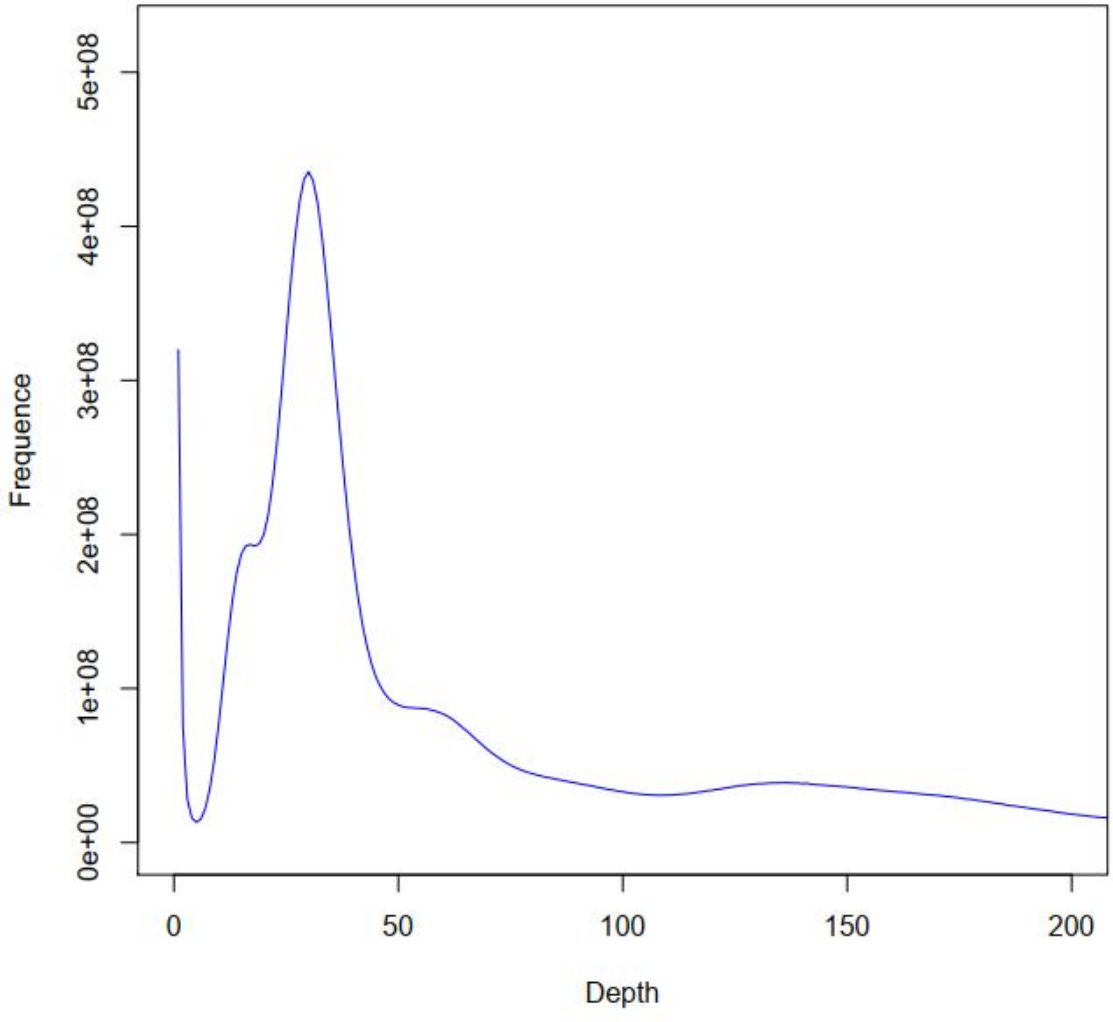

Fig. 3 17 K-mer analysis for estimating the genome size of *Monochamus alternatus* num

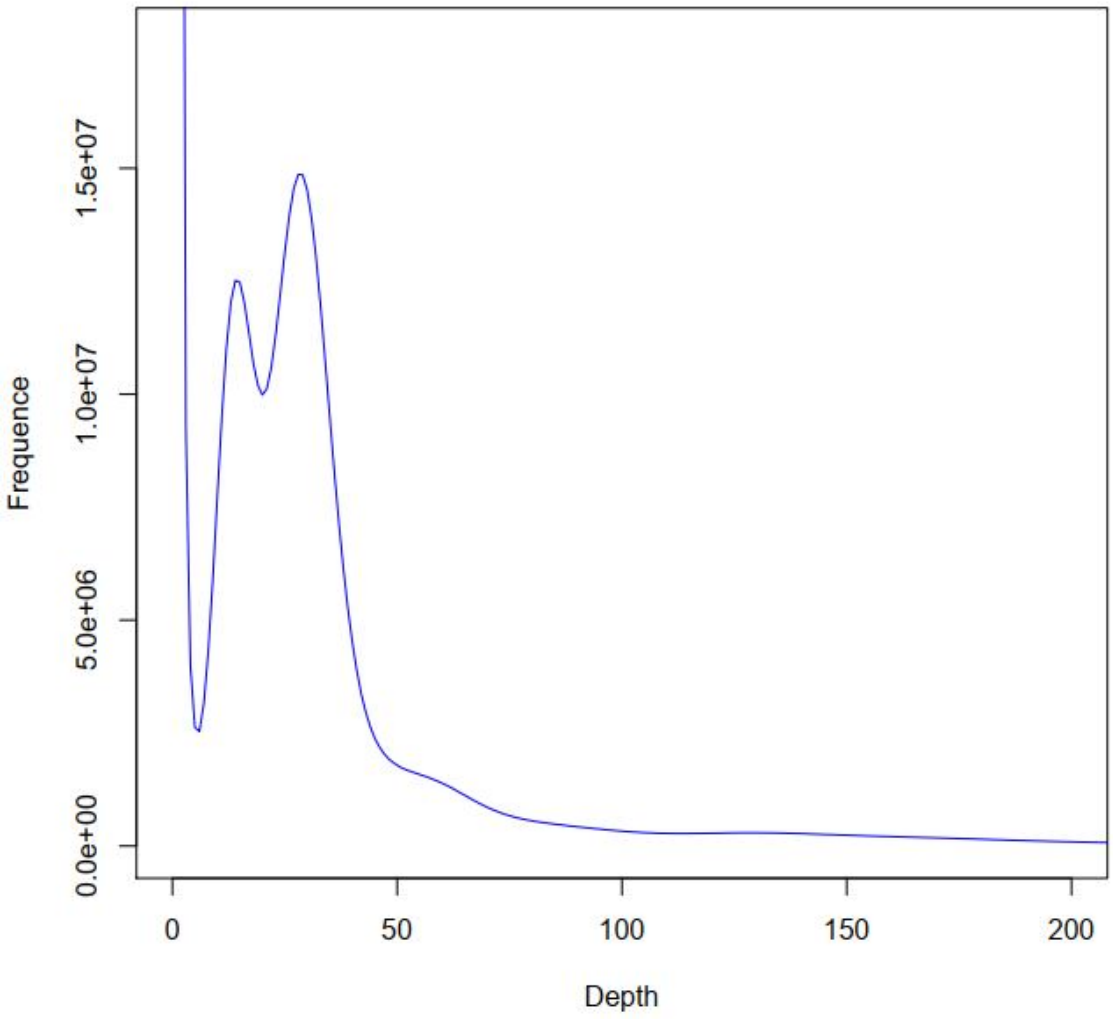

Fig.4 17 K-mer analysis for estimating the genome size of *Monochamus alternatus* spe

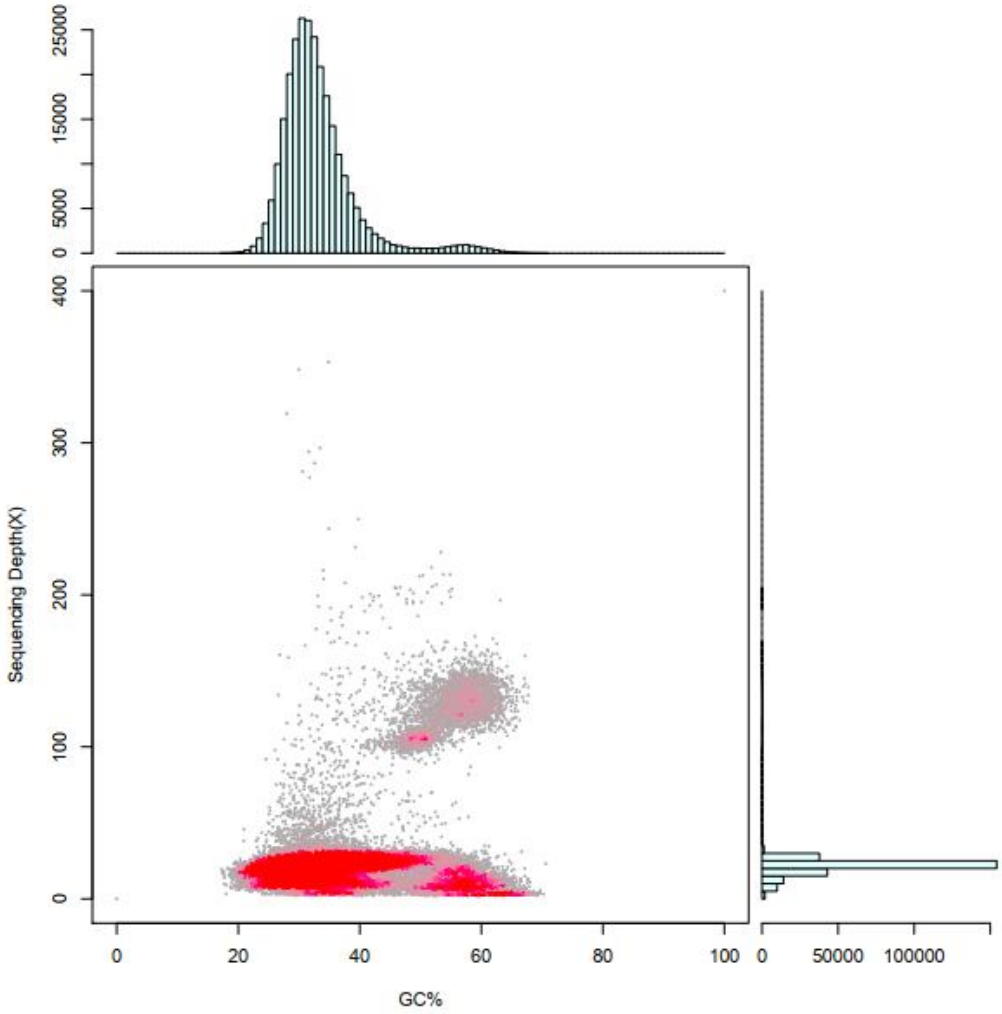

Fig. 5. GC content and average sequencing depth

Table 3. Genomic information statistics of *Monochamus alternatus*

Item	Conting	Scaffold
	Size (bp)	Size (bp)
N50	787	954
N90	135	145
Total Size	823 399 058	840 047 447
Total Number	2 034 890	1 839 509
Longest	385 511	458 244

Table 4. Genome assembly data of *Monochamus alternatus* and other species of Coleoptera

Species	Genome size (Mb)	Number of protein-coding genes	GC content (%)	Heterozygosity (%)	Repeat (%)	References
Monochamus alternatus	871.09	n.a.	32.94	1.04	59.04	This study
Diabrotica virgifera	2418.07	28061	36.5	n.a.	n.a.	Wu K,et al.,2017
Photinus pyralis	471.51	32294	36.4	0.598	42.6	Fallon TR et al.,2018
Listronotus bonariensis	1112.44	n.a.	31.3	0.18	70	Harrop TWR, et al.,2020
Propylea japonica	851.23	18018	35.13	0.9	58.22	Zhang L, et al.,2020
Marronus borbonicus	406.94	23278	35.9	0.2	29.2	Meyer JM, et al.,2016
Abscondita terminalis	499.65	20439	31.4	n.a.	n.a.	Chen X, et al.,2019
Lamprigera yunnana	1052.93	19438	34.1	n.a.	n.a.	Chen X, et al.,2019
Hycleus cichorii	99.17	13813	32.3	1.00	72.73	Wu YM, et al.，2018
Harmonia axyridis	466.69	n.a.	33.2	n.a.	n.a.	Ando T, et al.，2018
Hycleus phaleratus	90.87	13725	30.06	0.99	74.90	Du C, et al.,2017
Aethina tumida	234.34	17463	30.00	n.a.	n.a.	Evans JD, et al.,2018

Limonius californicus	1072.67	n.a.	35.00	0.21	n.a.	Andrews KR, et al.,2020
Oryctes borbonicus	406.20	8822	34.85	0.2	29.2	Meyer JM, et al.,2016
Nicrophorus vespilloides	195.27	18995	32.20	n.a.	n.a.	Cunningham CB, et al.,2015
Asbolus verrucosus	249.61	n.a.	32.90	n.a.	n.a.	Stanley Dean Rider Jr,2015
Protaetia brevitarsis	1143.92	n.a.	25.40	n.a.	72.29	Kim MJ, et al.,2014
Popillia japonica	531.53	n.a.	34.90	n.a.	n.a.	Yang W, et al.,2018
Anoplophora glabripennis	706.97	20632	33.40	n.a.	n.a.	Lowe TM, et al.,1997
Tenebrio molitor	280.78	n.a.	26.50	n.a.	n.a.	Liu LN, et al.,2014
Agrilus planipennis	353.07	22159	36.00	n.a.	n.a.	Lord NP, et al.,2016
Leptinotarsa decemlineata	641.99	19038	35.60	n.a.	n.a.	Sharkey CR, et al.,2017
Callosobruchus maculatus	1007.82	31345	37.7	n.a.	65.00	Sayadi A, et al.,2019
Priacma serrata	12.08	n.a.	35.90	n.a.	n.a.	Niehuis O, et al.,2012
Dendroctonus ponderosae	257.09	16791	38.45	n.a.	n.a.	Keeling CI, et al.,2016