Detailed classification of mariner-like elements, class II transposable elements, in six Apis species based on Dromar

Kakeru Yokoi[1,*], Kiyoshi Kimura[2], Hidemasa Bono[3,4]

1 Insect Genome Research and Engineering Unit, Division of Applied Genetics, Institute of Agrobiological Sciences (NIAS), National Agriculture and Food Research Organization (NARO), 1-2 Owashi, Tsukuba, Ibaraki 305-8634, Japan

2 Honeybee Research Group, Division of Animal Breeding and Reproduction, National Institute of Livestock and Grassland Science (NILGS), National Agriculture and Food Research Organization (NARO), Tsukuba, 2 Ikenodai, Tsukuba, Ibaraki 305-0901, Japan

3 Database Center for Life Science (DBCLS), Joint Support-Center for Data Science Research, Research Organization of Information and Systems, 1111 Yata, Mishima, Shizuoka 411-8540, Japan

4 Program of Biomedical Science, Graduate School of Integrated Sciences for Life, Hiroshima University, 3-10-23 Kagamiyama, Higashi-Hiroshima City, Hiroshima, 739-0046, Japan

*Corresponding author: Kakeru Yokoi

E-mail addresses:

Kakeru Yokoi: yokoi123@affrc.go.jp

Kiyoshi Kimura: kimura@affrc.go.jp

Hidemasa Bono: bonohu@hiroshima-u.ac.jp

Abstract

The transposable element (TE) *mariner* is a class II TE that moves within a host genome in a "copy and paste style" manner. *Mariner* and mariner-like elements (MLEs) are distributed in various species and their sequences are highly diverse; therefore, difficulties have been associated with the classification of these elements. A few TE in the genomes of *Api*s species mainly consist of *mariner* and MLE. Detailed classifications are needed to obtain a clearer understanding of *Apis* MLEs, particularly their movement and horizontal transfer. In the present study, MLEs were detected in the genomes of six *Apis* species using RepeatMasker. The MLEs were classified into 31 Dromar classes. In this classification, the MLEs in the *Apis mellifera* genome were classified into the 18 Dromar classes consisting of all five subfamilies, and the MLEs in other *Apis* species were classified into the Dromar classes from only *mauritiana* and *mellifera* subfamilies, suggesting the potential utilization of this classification in the analysis of TEs. A phylogenetic analysis was performed on each Dromar class, and the results obtained suggested that some *A. mellifera* MLEs invaded via horizontal transfer.

Introduction

Mariner is a class II transposable element (TE) that moves within a host genome in a "copy and paste style" manner (Oosumi et al., 1995; Plasterk et al., 1999; Robertson, 1995). *Mariner* was initially discovered in the *Drosophila mauritiana* genome (Jacobson et al., 1986). TE similar to *Mariner* were subsequently identified in the genomes of many plants, vertebrates, and invertebrates, and named mariner-like elements (MLEs). *Mariner* and MLE encode a transposase that possesses a D,D34D motif, which is characteristic of *mariner* and MLE (Plasterk et al., 1999). Another *mariner*-specific characteristic is the terminal inverted repeat (TIR) located at both ends of *mariner* that contains transposase-binding sites. Many MLEs and their subfamilies (e.g. *irritans* and *mellifera*) have been detected in various species using PCR methods with primers based on the sequences of TIR or the D,D34D motif (Yamada et al., 2014). Previous studies indicated that *mariner* and MLE distributed in the genomes of many species via horizontal transfer (Robertson, 1995). However, these PCR-based methods cannot detect all *mariner* and MLE in genomes because they amplify autonomous elements, but not some non-autonomous elements possessing several mutants or variations. MLEs which are changed into non-autonomous from autonomous cannot be traced. Due to the availability of whole genome sequence data and the lower cost of genome sequencing, the methods employed to identify TE, including *mariner* and MLE, have been modified. One method involves the use of "RepeatMasker" (Smit A.F.A., Hubley R. & Green P. RepeatMasker Open 4.0. 2013–2015, http://www.repeatmasker.org), which searches sequences in whole genome data that are similar to consensus TE sequences registered in a sequence library (e.g. Repbase) (Jurka et al., 2005). This method detects all types of TE and autonomous and non-autonomous elements in the whole genome region, thereby tracking MLEs and providing further insights into the horizontal transfer of MLE between species.

The genus *Apis* mainly consists of four species, and the following whole genome data are currently available (three *Apis* species and three *Apis cerana* subspecies): *A. mellifera* (Am) (Wallberg et al., 2019), *A. cerana japonica* (Ack) (Yokoi et al., 2018), *A. cerana cerana* korea strain (Ack) (Park et al., 2015), *A. cerana cerana* china strain (Acc) (Diao et al., 2018), *A. dorsata* (Ad) (Oppenheim et al., 2020), and *A. florea* (Af) (Table 1). *Apis* genomes contain small numbers of TE, which mainly consist of class II TE, particularly MLEs (Weinstock et al., 2006; Yokoi et al., 2018).

Difficulties have been associated with the classification of TE because of variations in their sequences, as described above. Six *mariner* consensus sequences (Ammers 1-6) have been reported for *A. mellifera* (Weinstock et al., 2006), and 36 Dromar classes for *Drosophila* species (Wallau et al., 2014). Detailed classifications, such as Dromar, may be applied to *Apis* MLEs. In

the present study, we extracted Dromar sequences in each Dromar class and built profile files of each Dromar classes by Clustal omega and hmmer with the Dromar sequences. *Apis* MLE were then classified using these profiles by nhmmer. Furthermore, a phylogenetic analysis of the *Apis* MLE of each Dromar class was performed (Fig .1). The results obtained provide novel insights into the horizontal transfer and movement of MLEs in the six *Apis* species examined.

Results

MLE in six *Apis* species

RepeatMasker was used to detect MLEs in the Am, Ack, Acc, Acj, Af, and Ad genomes. The output files obtained are shown in Supporting data 1 and 2. Approximately 50 to 150 MLE were detected in the Ack, Acc, Acj, Af, and Ad genomes, while approximately 2000 were identified in the Am genome. Detailed and sequence data of all detected MLE are shown in Supporting data 3 and 4, respectively.

Dromar profiling

To build profiling data on 36 Dromar groups, we extracted Dromar sequences using 20 *Drosophila* genome sequences and the positional data of Dromar reported by Wallau et al. (2014). We only obtained 1418 Dromar sequences from 3685 Dromar copies because some *Drosophila* genome data were not available in the public database or the descriptions of some scaffolds in *Drosophila* genome data differed from those in positional data (Supporting data 5). Therefore, sequences for Dromar3, 7, 18, 19, and 20 were not extracted. Consequently, the nhmmer profiles of 31 Dromars for the classification of *Apis* MLE were built from the alignment results of each Dromar group (separated and merged alignment and nhmmer profiling data are shown in Supporting data 6 and 7, respectively).

Apis MLE classification

Apis MLEs were classified using the extracted *Apis* MLE sequences and Dromar profiles. The results of the classification are shown in Fig. 2, and detailed classification results are provided in Supporting data 8. In total, 864 out of approximately 2000 Am MLE were classified into 18 Dromar classes. The majority of Am MLEs were classified into the *mauritiana* and *melliferra* subfamilies, while no MLE were classified into Dromar 24 or 25 (*mauritiana* subfamily). Moreover, 84, 11, and three Am MLEs were classified into the *irritans*, *vertumnnana*, and *drosophila* subfamilies, respectively, into which one or no MLE of the other five *Apis* species was classified. The results of the MLE classifications of the five other species

were not markedly different from each other (Fig. 2). The MLEs of the five species were not allocated into the *vertumnnana* or *drosophila* subfamily, while a single MLE from Af, Ad, and Acj was classified into Dromar35 (*irritans* subfamily). In the *mauritiana* subfamily, one or more MLEs were classified into Dromar1, 26, and 32. In some of the five species, no MLE was classified into Dromar9 or 27. There were no MLE of the five *Apis* species (plus Am) in Dromar24 or 25. All Dromars belonging to the *mellifera* subfamily had one or more MLE. More than ten MLEs from Acj, Ack, and Acc were classified into Dromar 17. A list of all classified MLEs is shown in Supporting data 8.

Phylogenetic analysis of *Apis* MLE and Dromar sequences.

Sequence alignments were performed using the Dromar and *Apis* MLE sequences belonging to each Dromar class, and a phylogenetic tree of each Dromar class was constructed based on alignment results (newick files for phylogenetic trees are shown in Supporting data 9). We selected a representative Dromar class for each subfamily. In Dromar26, one of the *mauritiana* subfamily, there were four distinct clades in the phylogenetic tree. Two clades consisted of only Am MLEs (Fig. 3A). The other clade included six *Apis* MLEs, but not Dromar sequences, while a clade included Dromar sequences and Am MLE sequences. In Dromar17, the *mellifera* subfamily, the phylogenetic tree consisted of five distinctive clades. The clade that was the most frequently isolated from the four other clades included only Am MLEs. The three clades consisted of the six *Apis* MLEs, while the other one only included Dromar sequences. In Dromar35, the *irritans* subfamily, two distinct clades were formed. One consisted of only Am MLEs and the other consisted of the four *Apis* MLEs and Dromars (Fig. 3C). In Dromar2, the *vertumnnana* subfamily, all Am MLE formed a single clade that markedly differed from the other Dromar clades (Fig. 3D). In Dromar31, the *drosophila* subfamily, three Am MLE formed a distinct cluster separate from Dromar sequences (Fig. 3E).

Discussion

In the present study, we extracted MLEs from six *Apis* genomes. MLEs were classified into Dromar classes according to hmm profiles made by Dromar sequences (Wallau et al., 2014). Some Am MLEs and six *Apis* MLEs in each Dromar class formed distinct clades in the phylogenetic tree.

More Am MLE were detected than the other five *Apis* MLE. Differences in the MLE numbers detected were not due to the reference genome qualities of the six *Apis* species. N50 of the *A. mellifera* genome (version Amel_4.0, the first version) was 40,524, and was similar to those of the Acc, Ack, Ad, and Ad genomes (Table 1). However, approximately 1100 MLE

were detected in the Amel_4.0 genome (Weinstock et al., 2006). These results revealed the presence of more MLE in the Am genome than in the other five *Apis* genomes.

The initial study on the Am genome reported that approximately 1100 MLEs were detected in the genome, and the MLEs have been classified into six classes, *AmMar1-6*, which include members of the *mellifera*, *irritans*, and *rosa* subfamilies (Weinstock et al., 2006). In the present study, we detected approximately 2000 MLEs from the latest version of the Am genome, and 864 MLEs of the detected MLEs with an e-value < 1e-5 were classified into 18 Dromar classes by nhmmer. The majority of the 864 MLE were allocated to the *mauritiana* and *mellifera* subfamilies, while the others were classified into the *irritans*, *vertumnnana*, and *drosophila* subfamilies. However, all MLEs of the five other *Apis* species, except for three (Acj, Ad, and Af MLE classified into Dromar 35), were classified into the *mauritiana* and *mellifera* subfamilies. No previous study in *Apis* species achieved such detailed classifications, especially classification into *mauritiana* subfamily, which has not been classified. This may be because nhmmer adopted a hidden markov model that detects previously unrecognized sequence features (Wheeler and Eddy, 2013). Therefore, our classification method using nhmmer may be employed in a TE analysis.

Phylogenetic trees revealed that there were clades including only Am MLEs or the six *Apis* MLEs, and most Am MLEs belonged to the former clades. TE may jump in a host genome, resulting in unfavorable or favorable phenotypes. In contrast to *A. cerana*, *A. florela*, and *A. dorsata*, *A. mellifera* is bred and obtained by human selection. *A. mellifera*-specific features may be associated with the Am MLEs forming the Am MLE only clades. Furthermore, some Am MLEs can be considered to be invaded by horizontal transfer and settled in *A. mellifera*. On the other hand, the MLEs of six *Apis* species forming a single clade may settle through *Apis* species evolution. Further detailed analyses of classification data will reveal the dynamics of some MLE within the six *Apis* species.

Conclusion

We detected MLEs from the genome data of six *Apis* species, and performed nhmmer-based *Apis* MLE classification and phylogenetic analyses. *Apis* MLEs were classified into Dromars, more detailed categories, which suggests that our nhmmer-based classification is applicable to the classification of TE. Phylogenetic analyses revealed that Am MLE formed distinct clades in each Dromar class, and the MLE of six *Apis* species also formed several other distinct clades. These results provide insights into the evolution and dynamics of MLE in *Apis* species.

Materials and Methods

MLE detection and extraction of sequence data

The six *Apis* genome sequences used in the present study are shown in Table 1. RepeatMasker (Smit A.F.A., Hubley R. & Green P. RepeatMasker Open 4.0. 2013–2015, http://www.repeatmasker.org) was used to detect TE in the six *Apis* genomes, as described by Yokoi et al. (2018). RepeatMasker .tbl and .out files are shown in Supporting data 1 and 2, respectively. Data on MLE were extracted from outfiles, and a single ID was allocated to each detected element (Supporting data 3). To extract each mariner sequence, positional data in Supporting data 3 and BLAST software (Altschul et al., 1990) was utilized (Supporting data 4).

Classification of *Apis* MLE

To build the hmmer profiles of Dromars, Dromar sequences were extracted using the *Drosophila* genome data and Dromar location data reported by Wallau et al. (2014) (Supporting data 5). However, some Dromar sequences were not extracted because genome sequence IDs in *Drosophila* genome data differed from those described in Dromar location data. Therefore, several Dromars were not used in the present study. The sequences of each Dromar were aligned by Clustal omega (Clustal Omega - 1.2.4) (Sievers et al., 2011), and the Hmmer profiles of Dromars were made by hmmer-building (HMMER 3.2.1) (Supporting data 6) (Wheeler and Eddy, 2013). Using the nhmmer search program with merged Dromar hmm profiles (Supporting data 7), the Apis MLEs similar to each Dromar profile were identified (e-values < 1e-5) (Wheeler and Eddy, 2013). The output data and table results of nhmmer are shown in Supporting data 10 and 11 respectively. Based on hmmer search results, each MLE was classified into hit Dromar. If a MLE was hit to multiple Dromars, it was classified into Dromar with the lowest e-value.

Phylogenetic analysis

Dromars and classified MLE were aligned by Clustal Omega-1.2.4, an approximately-maximum-likelihood phylogenetic analysis was performed using fasttree version 2.1.10 (Price et al., 2009) (original data are shown in Supporting data 8), and phylogenetic trees were drawn in MEGAX (Kumar et al., 2018).

Data availability

Analyzed data and additional results have been uploaded to figshare (DOI: 10.6084/m9.figshare.c.4894503).

References

Altschul, S.F., Gish, W., Miller, W., Myers, E.W., Lipman, D.J., 1990. Basic local alignment search tool. J. Mol. Biol. 215, 403–410. https://doi.org/10.1016/S0022-2836(05)80360-2

Diao, Q., Sun, L., Zheng, Huajun, Zeng, Z., Wang, S., Xu, S., Zheng, Huoqing, Chen, Y., Shi, Y., Wang, Y., Meng, F., Sang, Q., Cao, L., Liu, F., Zhu, Y., Li, W., Li, Z., Dai, C., Yang, M., Chen, S., Chen, R., Zhang, S., Evans, J.D., Huang, Q., Liu, J., Hu, F., Su, S., Wu, J., 2018. Genomic and transcriptomic analysis of the Asian honeybee Apis cerana provides novel insights into honeybee biology. Sci. Rep. 8, 1–14. https://doi.org/10.1038/s41598-017-17338-6

Jacobson, J.W., Medhora, M.M., Hartl, D.L., 1986. Molecular structure of a somatically unstable transposable element in Drosophila. Proc. Natl. Acad. Sci. 83, 8684–8688.

Jurka, J., Kapitonov, V.V., Pavlicek, A., Klonowski, P., Kohany, O., Walichiewicz, J., 2005. Repbase Update, a database of eukaryotic repetitive elements. Cytogenet. Genome Res. 110, 462–467.

Kumar, S., Stecher, G., Li, M., Knyaz, C., Tamura, K., 2018. MEGA X: Molecular Evolutionary Genetics Analysis across Computing Platforms. Mol. Biol. Evol. 35, 1547–1549. https://doi.org/10.1093/molbev/msy096

Lo, N., Gloag, R.S., Anderson, D.L., Oldroyd, B.P., 2010. A molecular phylogeny of the genus Apis suggests that the Giant Honey Bee of the Philippines, A. breviligula Maa, and the Plains Honey Bee of southern India, A. indica Fabricius, are valid species. Syst. Entomol. 35, 226–233. https://doi.org/10.1111/j.1365-3113.2009.00504.x

Oosumi, T., Belknap, W.R., Garlick, B., 1995. Mariner transposons in humans. Nature 378, 672–672.

Oppenheim, S., Cao, X., Rueppel, O., Krongdang, S., Phokasem, P., DeSalle, R., Goodwin, S., Xing, J., Chantawannakul, P., Rosenfeld, J.A., 2020. Whole Genome Sequencing and Assembly of the Asian Honey Bee Apis dorsata. Genome Biol. Evol. 12, 3677–3683.

Park, D., Jung, J.W., Choi, B.-S., Jayakodi, M., Lee, J., Lim, J., Yu, Y., Choi, Y.-S., Lee, M.-L., Park, Y., 2015. Uncovering the novel characteristics of Asian honey bee, Apis cerana, by whole genome sequencing. BMC Genomics 16, 1.

Plasterk, R.H., Izsvák, Z., Ivics, Z., 1999. Resident aliens: the Tc1/mariner superfamily of transposable elements. Trends Genet. 15, 326–332.

Price, M.N., Dehal, P.S., Arkin, A.P., 2009. FastTree: Computing Large Minimum Evolution Trees with Profiles instead of a Distance Matrix. Mol. Biol. Evol. 26, 1641–1650. https://doi.org/10.1093/molbev/msp077

Robertson, H.M., 1995. The Tcl-mariner superfamily of transposons in animals. J. Insect Physiol. 41, 99–105.

Sievers, F., Wilm, A., Dineen, D., Gibson, T.J., Karplus, K., Li, W., Lopez, R., McWilliam, H., Remmert, M., Söding, J., Thompson, J.D., Higgins, D.G., 2011. Fast, scalable generation of high-quality protein multiple sequence alignments using Clustal Omega. Mol. Syst. Biol. 7, 539. https://doi.org/10.1038/msb.2011.75

Wallau, G.L., Capy, P., Loreto, E., Hua-Van, A., 2014. Genomic landscape and evolutionary dynamics of mariner transposable elements within the Drosophila genus. BMC Genomics 15, 727.

Wallberg, A., Bunikis, I., Pettersson, O.V., Mosbech, M.-B., Childers, A.K., Evans, J.D., Mikheyev, A.S., Robertson, H.M., Robinson, G.E., Webster, M.T., 2019. A hybrid de novo genome assembly of the honeybee, Apis mellifera, with chromosome-length scaffolds. BMC Genomics 20, 275. https://doi.org/10.1186/s12864-019-5642-0

Weinstock, G.M. et al., The Honeybee Genome Sequencing Consortium, 2006. Insights into social insects from the genome of the honeybee Apis mellifera. Nature 443, 931–949. https://doi.org/10.1038/nature05260

Wheeler, T.J., Eddy, S.R., 2013. nhmmer: DNA homology search with profile HMMs. Bioinformatics 29, 2487–2489. https://doi.org/10.1093/bioinformatics/btt403

Yamada, K., Kawanishi, Y., Yamada, A., Tokuda, G., Gurung, R.D., Sasaki, T., Nakajima, Y., Maekawa, H., 2014. A novel cluster of mariner-like elements belonging to mellifera subfamily from spiders and insects: implications of recent horizontal transfer on the South-West Islands of Japan. Genetica 142, 149–160.

Yokoi, K., Uchiyama, H., Wakamiya, T., Yoshiyama, M., Takahashi, J.-I., Nomura, T., Furukawa, T., Yajima, S., Kimura, K., 2018. The draft genome sequence of the Japanese honey bee, Apis cerana japonica (Hymenoptera: Apidae). Eur. J. Entomol. 115.

Funding

This work was supported by ROIS-DS-JOINT (026RP2019 and 030RP2018) to KY and by the National Bioscience Database Center of the Japan Science and Technology Agency (JST) to HB.

Author contributions

Conceived and designed the experiments: K.Y., K.K., and H.B.
Analyzed the data: K.Y., K.K., and H.B.
Contributed to the writing of the manuscript: K.Y. and H.B.
All authors discussed the data and contributed to manuscript preparation. K.Y. supervised the project.
All authors read and approved the final manuscript.

Conflicts of interest

The authors declare no conflict of interest. The funders had no role in the design of the study; in the collection, analyses, or interpretation of data; in the writing of the manuscript, or in the decision to publish the results.

Figures & Tables

Fig. 1 Scheme of data analysis in the present study

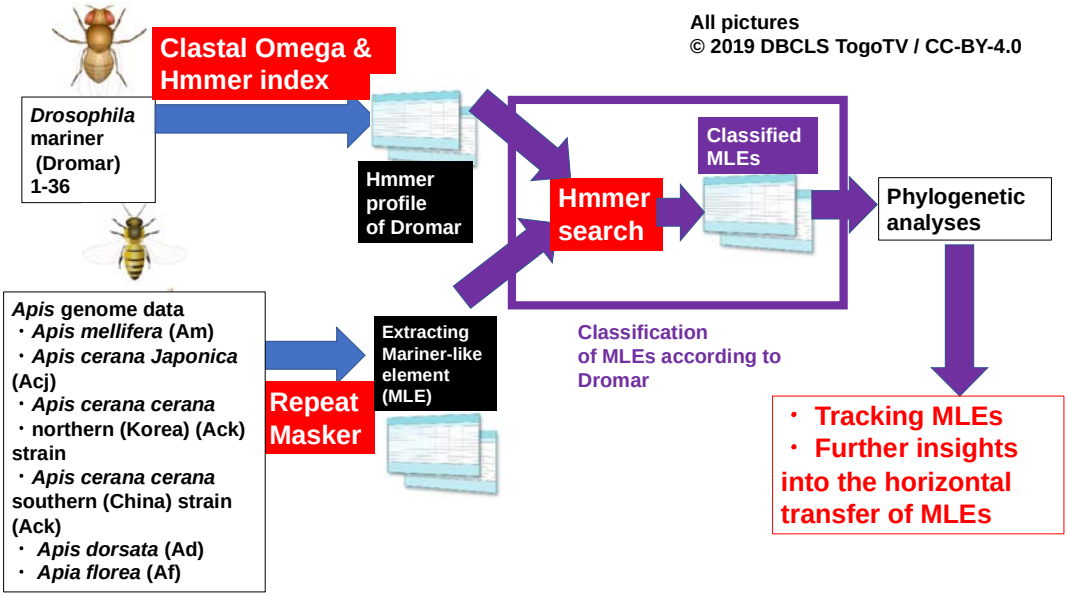


Fig. 2 Apis MLE allocated to Dromars

Number of *Apis* MLE allocated to each Dromar (e-value <1e-5). The left drawing shows the molecular phylogeny of the genus *Apis* based on the study by Lo et al. (2010). The color of each column indicates the subfamilies of *mariner* to which Dromar belongs, based on Wallau et al. (2014). Unnumbered Dromars are those whose sequences could not be extracted from the genome. Detailed classification lists are shown in Supporting data 8.

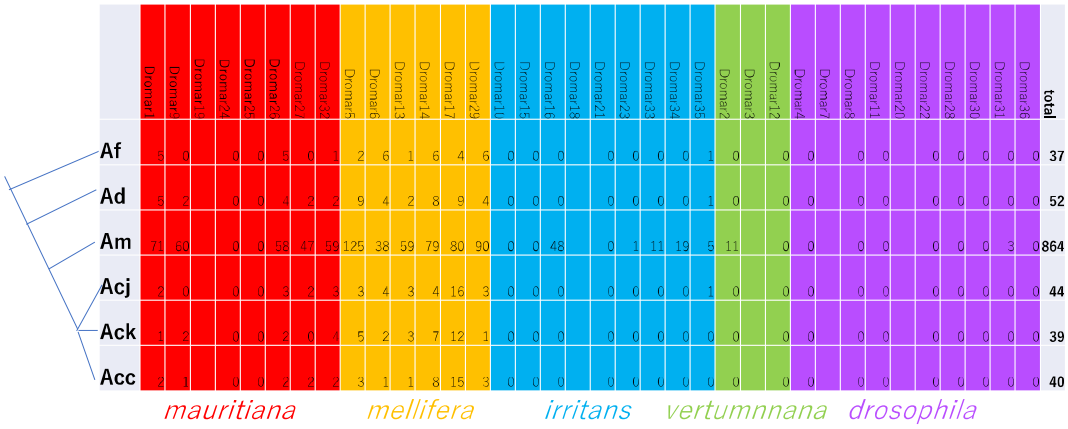
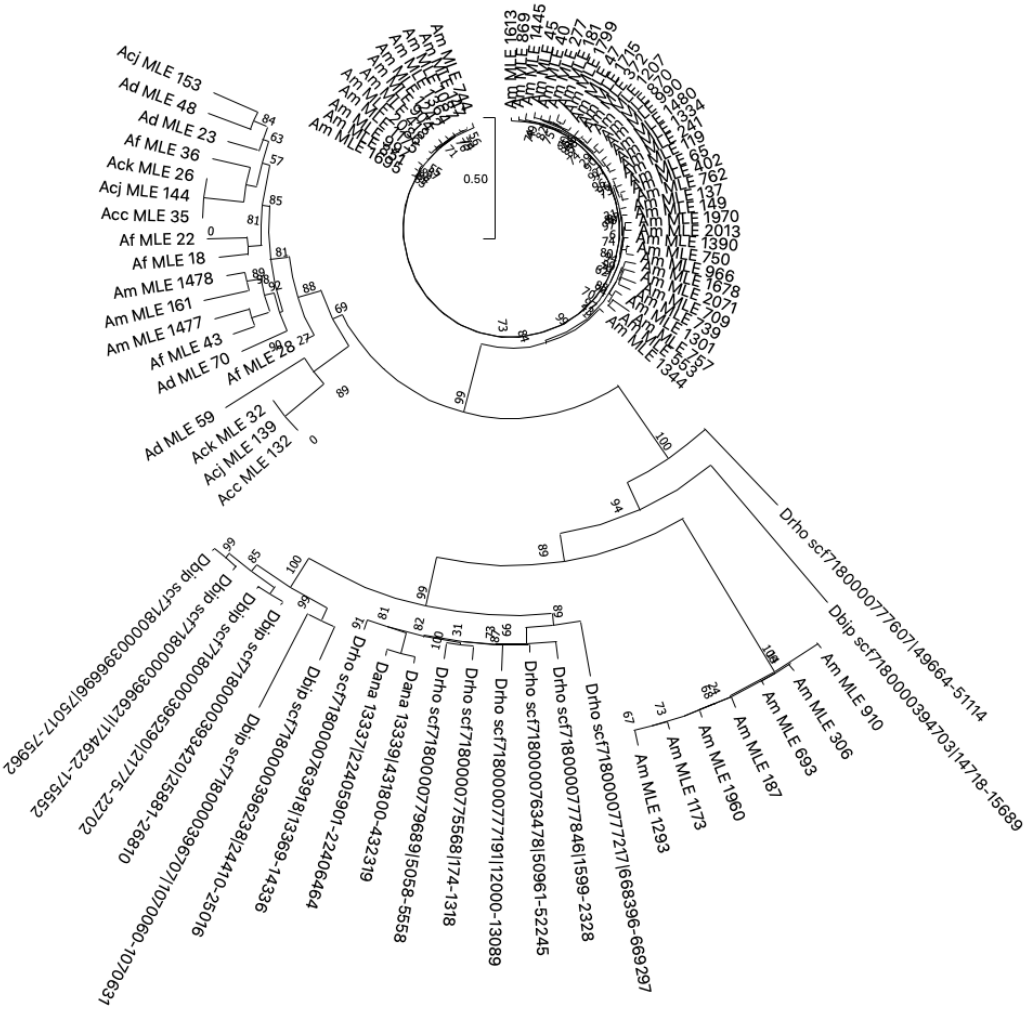


Fig. 3 Representative phylogenetic tree of five *mariner* subfamilies

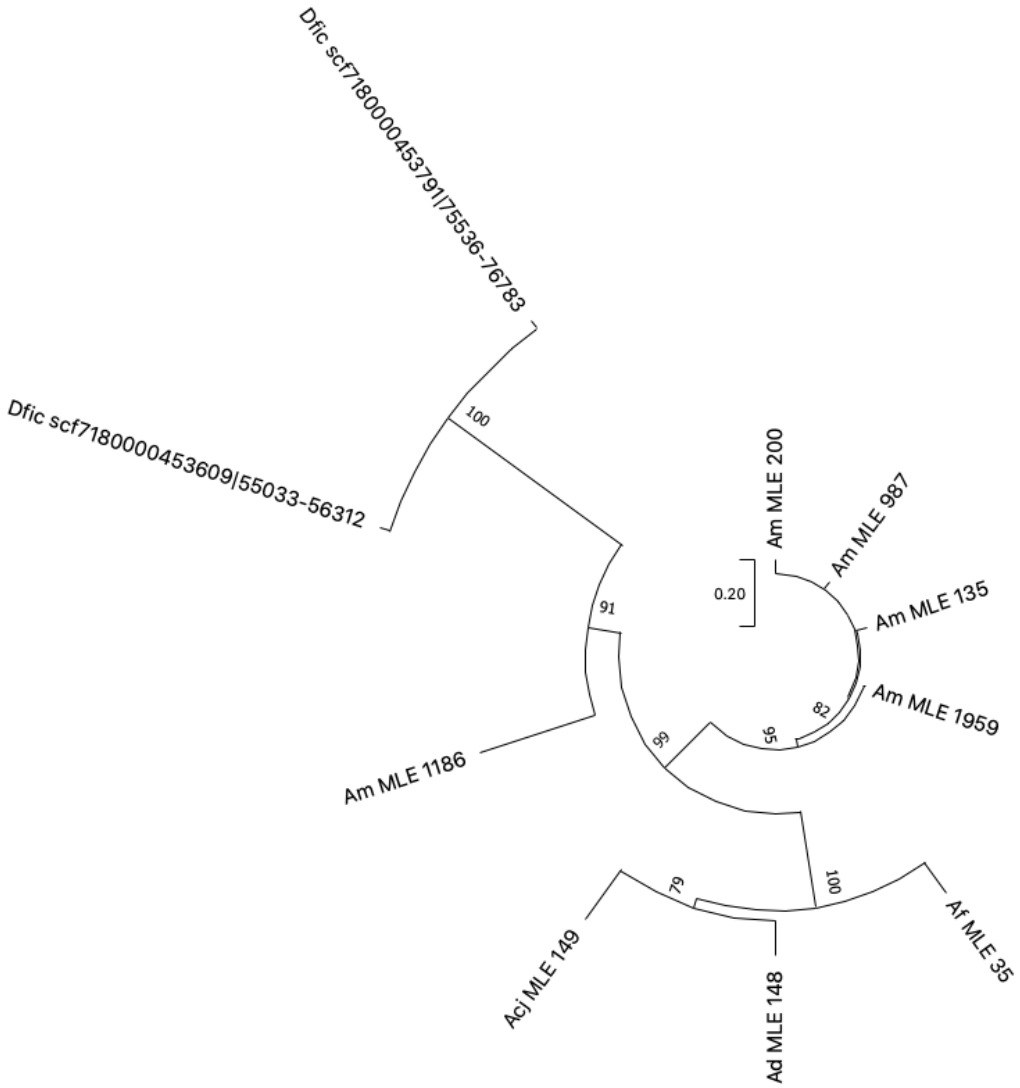
Phylogenetic trees based on the alignments of *Apis* MLE and Dromar sequences from representative Dromar classes of the five subfamilies: *mauritiana* (A), *mellifera* (B), *vertumnnana* (C), *irritans* (D), and *drosophila* (E).

(A) Dromar26 (*mauritiana*)

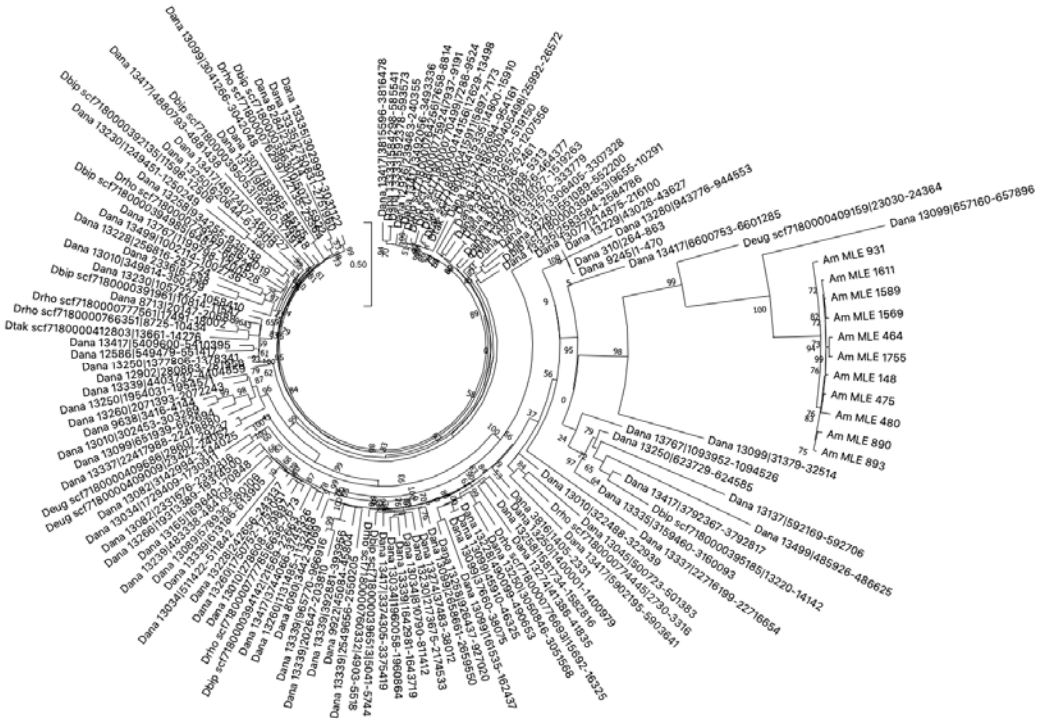


(B) Dromar17 (*mellifera*)

(C) Dromar35 (*irritans*)



(D) Dromar2 (*vertumnnana*)


Dana 13099|657160-657896
Deug scf7180000409159|23030-24364
Am MLE 931
Am MLE 1611
Am MLE 1589
Am MLE 1569
Am MLE 464
Am MLE 1755
Am MLE 148
Am MLE 475
Am MLE 480
Am MLE 890
Am MLE 893
Dana 13767|1093952-1094526
Dana 13250|623729-624585
Dana 13417|3792367-3792817
Dana 13137|592169-592706
Dana 13499|485926-486625
Dana 13335|3159460-3160093
Dana 13337|2276169-2276654
Dana 13010|322488-323939
Dana 13250|1400001-1400979
Dana 13417|6902195-5903641
Dana 13417|6600753-6601285
Dana 13280|943776-944553
Dana 13229|43028-43627
Dana 13229|214575-216100
Dana 310|264-863
Dana 924|1-470
Dana 13417|3815596-3816478
Dana 13417|3374305-3375419
Dana 13339|3922445-3923205
Dana 13339|3922445-3923205
Dana 13339|3922445-3923205

(E) Dromar31 (*drosophila*)

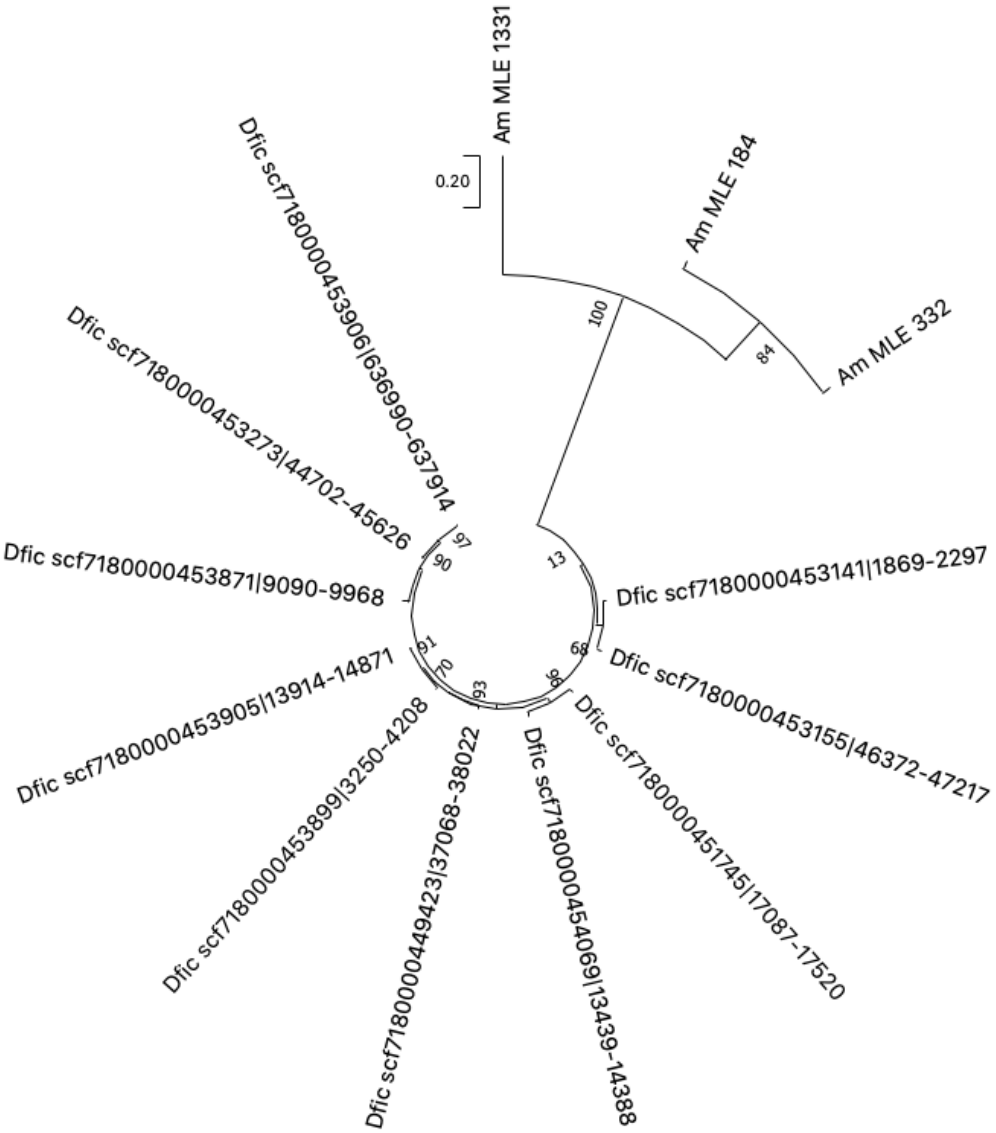


Table 1 *Apis* genome assemblies used in the present study

Organism name	GenBank assembly accession	Contig N50
A. mellifera	GCA_003254395.2	5,382,476
A. cerana japonica	GCA_002217905.1	179,487
A. cerana cerana Korea	GCA_001442555.1	43,751
A. cerana cerana China	GCA_002290385.1	21,160
A. dorsata	GCA_009792835.1	30,868
A. florea	GCA_000184785.2	24,915

Supporting data

All Supporting data have been uploaded in figshare (DOI:10.6084/m9.figshare.c.4894503)

Supporting data 1

Out files of RepeatMasker from six *Apis* genomes. DOI:10.6084/m9.figshare.11914278

Supporting data 2

Tbl files of RepeatMasker from six *Apis* genomes. DOI: 10.6084/m9.figshare.11984451

Supporting data 3

MLE data on six *Apis* species. DOI: 10.6084/m9.figshare.11914263

Supporting data 4

MLE sequences. DOI: 10.6084/m9.figshare.11913309

Supporting data 5

Dromar sequences (Wallau et al., 2014) used in the present study.
DOI:10.6084/m9.figshare.11956350

Supporting data 6

Alignment and hmmer profile files of Dromars. DOI: 10.6084/m9.figshare.11962614

Supporting data 7

Merged Dromar hmmer profile for MLE classifications. DOI: 10.6084/m9.figshare.11953734

Supporting data 8

List of *Apis* MLEs and classification results DOI:10.6084/m9.figshare.12045189

Supporting data 9

Original data of the phylogenetic analysis in newick files. DOI:10.6084/m9.figshare.11955621

Supporting data 10

Results of nhmmer of *Apis* MLEs against Dromar profiles. DOI:10.6084/m9.figshare.11953746

Supporting data 11

Result tables of nhmmer of *Apis* MLEs against Dromar profiles. DOI: 10.6084/m9.figshare.11953752