ORPA: A Fast and Efficient Method for Constructing Genome-Wide Alignments of Organelle Genomes for Phylogenetic Analysis

Guiqi Bi[1,#*], Xinxin Luan[1,2#] Jianbin Yan[1*]

[1]*Shenzhen Branch, Guangdong Laboratory of Lingnan Modern Agriculture, Key Laboratory of Synthetic Biology, Ministry of Agriculture and Rural Affairs, Agricultural Genomics Institute at Shenzhen, Chinese Academy of Agricultural Sciences, Shenzhen 518124, China*

[2]*Zhengzhou Research Base, State Key Laboratory of Cotton Biology, School of Agricultural Sciences, Zhengzhou University, Zhengzhou, China*

[#] These authors contributed equally to this work.

[*] Correspondence to: biguiqi@caas.cn (G.B.); jianbinlab@caas.cn (J.Y.)

SUMMARY

Creating a multi-gene alignment matrix for phylogenetic analysis using organelle genomes involves aligning single-gene datasets manually, a process that can be time-consuming and prone to errors. The HomBlocks pipeline has been created to eliminate the inaccuracies arising from manual operations. The processing of a large number of sequences, however, remains a time-consuming task. To conquer this challenge, we have developed a speedy and efficient method called ORPA. ORPA quickly generates multiple sequence alignments for whole-genome comparisons by parsing the result files of NCBI BLAST, completing the task in just one minute. With increasing data volume, ORPA's efficiency is even more pronounced, over 300 times faster than HomBlocks in aligning 60 high-plant chloroplast genomes. The tool's phylogenetic tree outputs demonstrate equivalent results to HomBlocks, indicating its outstanding efficiency. Due to its speed and accuracy, ORPA can identify species-level evolutionary conflicts, providing valuable insights into evolutionary cognition.

KEYWORDS

Organelle phylogenomics, Phylogenomic conflict, Efficient pipeline

INTRODUCTION

Phylogenetic trees utilizing organelle genomes are becoming indispensable in comparative genomics and systematics. They play a crucial role in elucidating the evolutionary relationships among species, particularly when incomplete lineage sorting obscures these relationships. This approach provides a broad perspective and facilitates a more accurate assessment of the phylogenetic relationships among species. It is a valuable tool for studying the complex evolutionary history of eukaryotic life (Li et al., 2019; Li et al., 2021).

Creating a precise multiple sequence alignment (MSA) is critical to constructing an accurate phylogenetic tree. This typically involves aligning single-copy genes beforehand and then concatenating them to form a super matrix. However, this can be laborious and error-prone, often requiring manual adjustments. To tackle this challenge, we previously introduced HomBlocks software (Bi et al., 2018), which enhances the accuracy and efficiency of MSA construction for organelle genome sequences by eliminating the need for manual operations. However, when dealing with a substantial number of sequences, HomBlocks may still struggle with processing speed, which remains unsatisfactory. To address this issue more effectively, we have developed an innovative approach called ORPA. This method ranks as the fastest software for constructing multi-sequence alignments (MSA) of organelle genomes and delivers exceptionally accurate results compared to HomBlocks. Our research findings offer compelling evidence that ORPA is a highly viable option for HomBlocks, ensuring superior speed and accuracy. Moreover, the ORPA can be employed in systematic investigations to promptly obtain precise evolutionary relationships among species, resulting in significant research discoveries such as species-level evolutionary conflicts.

RESULTS

Comparison of tree topologies constructed from ORPA and HomBlocks sequence alignment

We evaluated the topological structures of phylogenetic trees constructed using ORPA and HomBlocks by employing the benchmark data used in the HomBlocks publication (Bi et al., 2018). As a direct comparison of the MSA results was impossible, we used this approach to observe similarities and differences between the two methods. The construction of phylogenetic trees was carried out using two approaches, the maximum likelihood method (RAxML (Stamatakis, 2014)) and Bayesian method (MrBayes 3.2.5 (Ronquist et al., 2012)).

In the first test dataset, we evaluated the performance of two software programs using a test dataset consisting of chloroplast genomes from 52 higher plants (Supplementary Table 1), as obtained from Zhang et al. (Zhang et al., 2016). Our results showed that the alignment lengths generated by ORPA and HomBlocks were 90,925 bp and 62,101 bp, respectively, with 8,270 bp and 8,404 bp of parsimony-informative sites. The resulting phylogenetic trees constructed from these two approaches exhibited identical topological structures with high support for all nodes except five (Fig. 2). For our second test dataset, we employed 36 mitochondrial genomes from Xenarthrans (Gibb et al., 2015) (Supplementary Table 2) to construct phylogenetic trees. Our comparative analysis of the resulting tree structures indicates that at non-100% support nodes, the support values from ORPA were slightly lower than those by HomBlocks. However, both methods exhibited a high degree of consistency in the overall topology of the phylogenetic tree (Fig. 3). These findings demonstrate the effectiveness of both approaches in generating reliable phylogenetic trees. Notably, all Multiple Sequence Alignment (MSA) constructions were accomplished within a time frame of less than 5 minutes using ORPA.

Moreover, we evaluated a set of higher plant mitochondrial datasets consisting of 18 publicly accessible mitochondrial sequences (Supplementary Table 3). Comparative assessments of the resulting phylogenetic trees corroborated the dependability of the ORPA and HomBlocks software suites (Fig. 3a). Additionally, we showcased the distribution of multiple sequence alignment (MSA) derived from both tools aligning to a reference genome, and the findings highlighted a remarkable concordance between the two methods, except for a few dissimilarities at specific loci (Fig. 3b).

Comparing the Runtime of ORPA and HomBlocks

To directly compare the runtime differences of ORPA and HomBlocks on the same dataset, we tested 60 higher plant chloroplast genomes (Supplementary Table 4). First, we constructed genome-wide alignments using ORPA based on these 60 chloroplast sequences and conducted Maximum likelihood tree reconstruction using IQ-TREE (Nguyen et al., 2014). Then, we sampled based on the topology of the tree and compared the runtime of ORPA and HomBlocks. The sampling range increased by 5 with each deepening of the evolutionary relationship (Fig. 5a). We ran ORPA and HomBlocks for each dataset and calculated their respective running times in minutes. Since ORPA typically runs quickly, we standardized the comparison to 1 minute. Additionally, we used the alignment results from each dataset to construct a fast ML tree using IQ-TREE and compared the resulting trees generated by the two tools. To display the differences in running time, we presented a bar chart (Fig. 5b). For the comparison of the resulting trees from each dataset, we expressed the results as a similarity percentage (Fig. 5b).

Figure 5b shows a significant increase in HomBlocks' runtime beyond 30 sequences, taking 313 minutes to complete the alignment process with 60 sequences. In contrast, based on

the same data source, ORPA can process faster than HomBlocks when the number of sequences exceeds 10. Additionally, the runtime of HomBlocks increases exponentially with the number of sequences, which could become more significant with datasets containing over 60 sequences. Conversely, the script runtime for ORPA remains unaffected by the number of sequences, except during the data preparation stage. This is mainly attributed to the utilization of BLAST as the kernel for the alignment process, which avoids the need for single-threaded sequence comparison.

This highlights ORPA's advantage over HomBlocks when dealing with a large number of sequences. Additionally, we conducted tree reconstructions for each pairwise test dataset and compared the ML tree topologies generated by ORPA and HomBlocks using treedist (https://github.com/agormp/treedist). Except for the comparison group with a sampling range of 25, the similarity between the tree topologies is 91%, indicating almost identical phylogenetic tree topologies generated by ORPA and HomBlocks in the other 11 comparison groups. Our findings suggest that ORPA outperforms HomBlocks in terms of speed and accuracy, offering researchers a powerful tool for creating whole-genome alignments.

Using ORPA for rapid detection of systematic evolutionary conflicts

Advancements in sequencing technology have led to the accumulation of a vast amount of organellar genome data. Effective utilization of this data has become a growing field of interest. This is especially important for newly sequenced data, as rapid confirmation of species' evolutionary relationships is crucial to verify sequencing accuracy. Additionally, constructing phylogenetic trees with speed and accuracy to investigate evolutionary conflicts is a key area of research in systematics biology. ORPA offers an elegant approach to achieving these goals. To provide a more comprehensive demonstration, we utilized 52 Lamiales chloroplast datasets as an example (Supplementary Table 5). We employed ORPA to build a multiple sequence alignment

of 101,454 characters, and constructed a corresponding phylogenetic tree to depict the evolutionary relationships. Figure 5 demonstrates that there are two apparent conflicts between the phylogenetic branches in regards to *Wightia speciosissima* and *Comoranthus minor*.

Wightia speciosissima, an angiosperm, has been assigned to a distinct family (Wightiaceae) by the Angiosperm Phylogeny Group IV (The Angiosperm Phylogeny, 2016). Its previous classification placed it within the Paulowniaceae family. However, analysis of its evolutionary branching suggests that it shares closer evolutionary relationships with the Phrymaceae family, thus representing a distinct lineage. This observation was also made by Xia et al (Xia et al., 2019). Their study on plant phylogeny utilized data from nine chloroplast genes and one mitochondrial *rps*3 gene. Consequently, they advise against including *Wightia speciosissima* in the Paulowniaceae family and suggest that it may instead be a hybrid origin between early lineages of Phrymaceae and Paulowniaceae (Xia et al., 2019).

This phylogenetic tree also reveals the incongruity between the genus *Comoranthus* and *Schrebera* in terms of their phylogenetic relationships. *Schrebera* is found in Africa and India, while *Comoranthus* is only found in Madagascar and the Comoros Islands (Wallander et al., 2000). Both species have similar fruit morphology: capsules with a woody ovary, ruffled epidermis, and split in half when mature. They contain seeds and are suborbicular and pear-shaped (Engel, 1968). Our analysis of the evolutionary relationships among these studied species reveals a paraphyletic relationship, with *Comoranthus* found nested within *Schrebera*. This outcome is consistent with recent findings by Hong-Wa et al., who suggest the genera should be synonymized (Hong-Wa et al., 2023). Incorporating this finding into taxonomic classification will aid in a more accurate understanding of the evolutionary history of these plant groups.

In summary, ORPA has demonstrated considerable promise in the realm of systematic taxonomy, as demonstrated by the aforementioned use cases.

EXPERIMENTAL PROCEDURES

Methods

The framework of ORPA is written in Perl. As tools for aligning genomes, HomBlocks uses a method of identifying locally collinear blocks (LCBs), while the main difference with ORPA is its strategy of directly parsing the NCBI BLAST online tool results. By avoiding the need for software installation and various dependencies, this approach simplifies genome alignment for novices in the field of bioinformatics (Fig. 1). The core of ORPA is based on the widely-used BLAST tool (McGinnis et al., 2004), which offers significant improvements in the efficiency and speed of sequence alignments. Compared to HomBlocks, ORPA is able to construct alignment files within 5 minutes on average. In contrast, HomBlocks requires an increasing amount of processing time as the number of sequences being aligned grows due to the single-threaded operation of its core software, Mavue (Darling et al., 2004). Therefore, ORPA offers a more efficient and versatile alternative to HomBlocks.

ORPA also provides users with four trimming methods, namely Gblocks (Castresana, 2000), trimAl (Capella□Gutiérrez et al., 2009), Noisy (Dress et al., 2008), and BMGE (Criscuolo et al., 2010), which are same to those offered by HomBlocks. Importantly, users can directly use the output results from ORPA to facilitate the construction of a phylogenetic tree, thus streamlining the sequence alignment process.

Implementation

ORPA is a rapid tool for constructing multiple sequence alignments of organelles. It is a command-line tool and functional under any version of Linux without the need for external installation. The Perl source code of ORPA is freely available for download at https://github.com/BGQ/ORPA.git, and comprehensive documentation and tutorials can be found at https://github.com/BGQ/ORPA.git.

ACKNOWLEDGMENTS

We sincerely thank the editors and reviewers for their valuable suggestions and comments on this study. This work was supported by the National Key R&D Program of China (2018YFA0903200), Science Technology and Innovation Commission of Shenzhen Municipality of China (ZDSYS 20200811142605017). It was also supported by the Elite Young Scientists Program of CAAS and the Agricultural Science and Technology Innovation Program.

AUTHOR CONTRIBUTIONS

GQB and JBY conceived and designed the study. GQB and XXL collected data. GQB and XXL performed data analysis. GQB and JBY wrote the manuscript with other authors providing recommendations for modifications.

CONFLICT OF INTEREST

The authors declare that they have no conflicts of interest associated with this work.

DATA AVAILABILITY STATEMENT

The attached file contains a list of example data and organelle genome sequences that are referred to in the main body of this study.

SUPPORTING INFORMATION

Additional Supporting Information may be found in the online version of this article.

Table S1. Accession numbers of 52 higher plant chloroplast genomes from GenBank used in the first example dataset.

Table S2. Accession numbers of 36 xenarthrans mitochondrial genomes from GenBank used in the second example dataset.

Table S3. Accession numbers of 18 higher plant mitochondrial genomes from GenBank used in the third example dataset.

Table S4. Accession numbers of 60 higher plant chloroplast genomes from GenBank used in the fourth example dataset.

Table S5. Accession numbers of Lamiales chloroplast genomes from GenBank used in the fifth example dataset.

REFERENCES

Bi, G., Mao, Y., Xing, Q. & Cao, M. (2018) HomBlocks: a multiple-alignment construction pipeline for organelle phylogenomics based on locally collinear block searching. *Genomics*, **110**, 18-22.

Capella□Gutiérrez, S., Silla□Martínez, J.M. & Gabaldón, T. (2009) trimAl: a tool for automated alignment trimming in large-scale phylogenetic analyses. *Bioinformatics*, **25**, 1972-1973.

Castresana, J. (2000) Selection of Conserved Blocks from Multiple Alignments for Their Use in Phylogenetic Analysis. *Molecular Biology and Evolution*, **17**, 540-552.

Criscuolo, A. & Gribaldo, S. (2010) BMGE (Block Mapping and Gathering with Entropy): a new software for selection of phylogenetic informative regions from multiple sequence alignments. *BMC Evolutionary Biology*, **10**, 210.

Darling, A.C.E., Mau, B., Blattner, F.R. & Perna, N.T. (2004) Mauve: Multiple Alignment of Conserved Genomic Sequence With Rearrangements. *Genome Research*, **14**, 1394-1403.

Dress, A.W., Flamm, C., Fritzsch, G., Grunewald, S., Kruspe, M., Prohaska, S.J. *et al.* (2008) Noisy: identification of problematic columns in multiple sequence alignments. *Algorithms for Molecular Biology*, **3**, 7.

Engel, P.P. (1968) The induction of biochemical and morphological mutants in the moss Physcomitrella patens. *American Journal of Botany*, **55**, 438-446.

Gibb, G.C., Condamine, F.L., Kuch, M., Enk, J., Moraes□Barros, N., Superina, M. *et al.* (2015) Shotgun Mitogenomics Provides a Reference Phylogenetic Framework and Timescale for Living Xenarthrans. *Molecular Biology and Evolution*, **33**, 621-642.

Hong-Wa, C., Dupin, J., Frasier, C., Schatz, G. & Besnard, G. (2023) Systematics and biogeography of Oleaceae subtribe Schreberinae, with recircumscription and revision of its Malagasy members. *Botanical Journal of the Linnean Society*.

Li, H.T., Luo, Y., Gan, L., Ma, P.F., Gao, L.M., Yang, J.B. *et al.* (2021) Plastid phylogenomic insights into relationships of all flowering plant families. *BMC Biology*, **19**, 1-13.

Li, H.T., Yi, T.S., Gao, L.M., Ma, P.F., Zhang, T., Yang, J.B. *et al.* (2019) Origin of angiosperms and the puzzle of the Jurassic gap. *Nature Plants*, **5**, 461-470.

Mcginnis, S. & Madden, T.L. (2004) BLAST: at the core of a powerful and diverse set of sequence analysis tools. *Nucleic Acids Research*, **32**, W20-W25.

Nguyen, L.T., Schmidt, H.A., Von Haeseler, A. & Minh, B.Q. (2014) IQ-TREE: A Fast and Effective Stochastic Algorithm for Estimating Maximum-Likelihood Phylogenies. *Molecular Biology and Evolution*, **32**, 268-274.

Ronquist, F., Teslenko, M., Van Der Mark, P., Ayres, D.L., Darling, A., Höhna, S. *et al.* (2012) MrBayes 3.2: Efficient Bayesian Phylogenetic Inference and Model Choice Across a Large Model Space. *Systematic Biology*, **61**, 539-542.

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

The Angiosperm Phylogeny, G. (2016) An update of the Angiosperm Phylogeny Group classification for the orders and families of flowering plants: APG IV. *Botanical Journal of the Linnean Society*, **181**, 1-20.

Wallander, E. & Albert, V.A. (2000) Phylogeny and classification of Oleaceae based on rps16 and trnL-F sequence data. *American Journal of Botany*, **87**, 1827-1841.

Xia, Z., Wen, J. & Gao, Z. (2019) Does the Enigmatic Wightia Belong to Paulowniaceae (Lamiales)? *Frontiers in Plant Science*, **10**, 528.

Zhang, D., Li, K., Gao, J., Liu, Y. & Gao, L.-Z. (2016) The Complete Plastid Genome Sequence of the Wild Rice Zizania latifolia and Comparative Chloroplast Genomics of the Rice Tribe Oryzeae, Poaceae. *Frontiers in Ecology and Evolution*, **4**, 88.

FIGURE LEGENDS

Fig. 1 Comparison of ORPA and HomBlocks efficiency in the conventional workflow for the phylogenetic tree construction of organelle genomes.

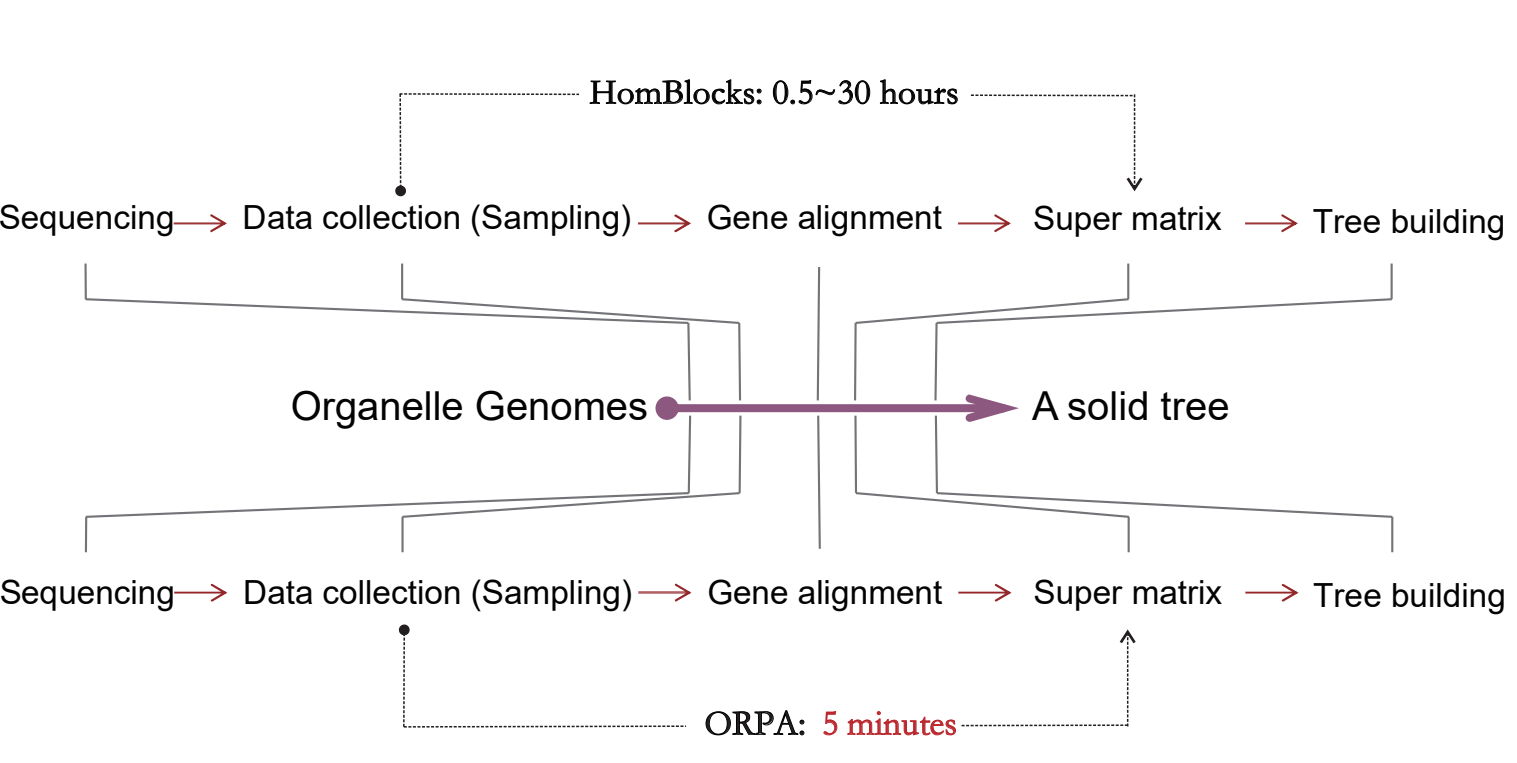
Fig. 2 Comparison of topology between the HomBlocks tree (left) and the ORPA tree (right) of 52 higher plant chloroplast genomes. The phylogenetic trees were constructed using maximum likelihood (ML) and Bayesian inference (BI) methods with the HomBlocks alignment (62,101 characters) and the ORPA alignment (90,925 characters), respectively. The support values inferred from RAxML (left) and Bayesian posterior probability (right) are indicated by the numbers on the nodes. Fully resolved nodes are not labeled with numbers. These results provide insights into the comparative performance of the two alignment methods for phylogenetic analysis of chloroplast genomes in higher plants.

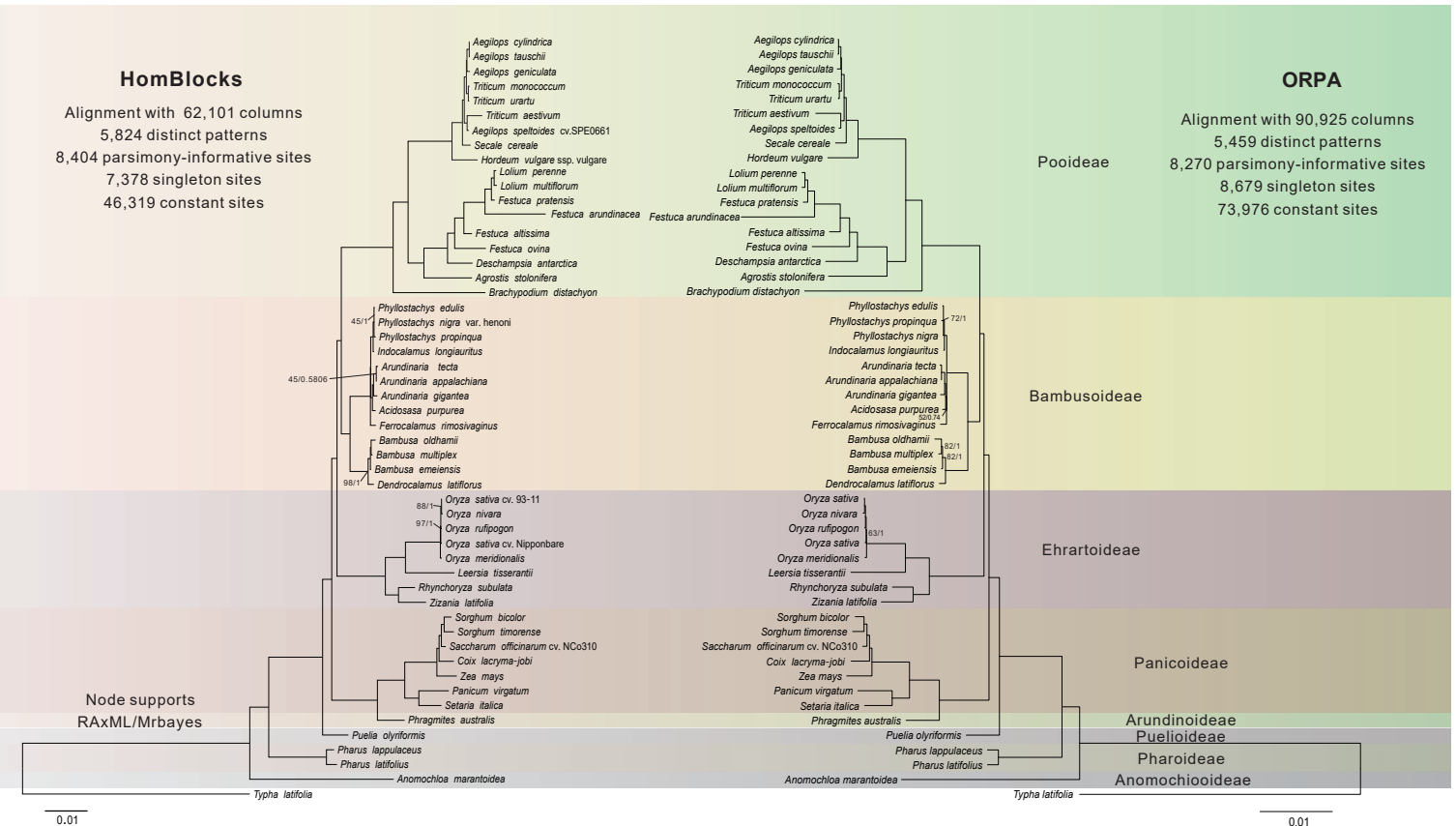
Fig. 3 Topology comparison of two phylogenetic trees of 36 xenarthran mitochondrial genomes. The HomBlocks tree (left) and the ORPA tree (right) were constructed using different alignment methods, one with 15,170 characters and the other with 8,696 characters. Maximum likelihood and Bayesian inference methods were used to construct the trees, and the support values derived from RAxML (left) and Bayesian posterior probability (right) are indicated by the numerical values on the nodes. Fully resolved nodes are unlabeled.

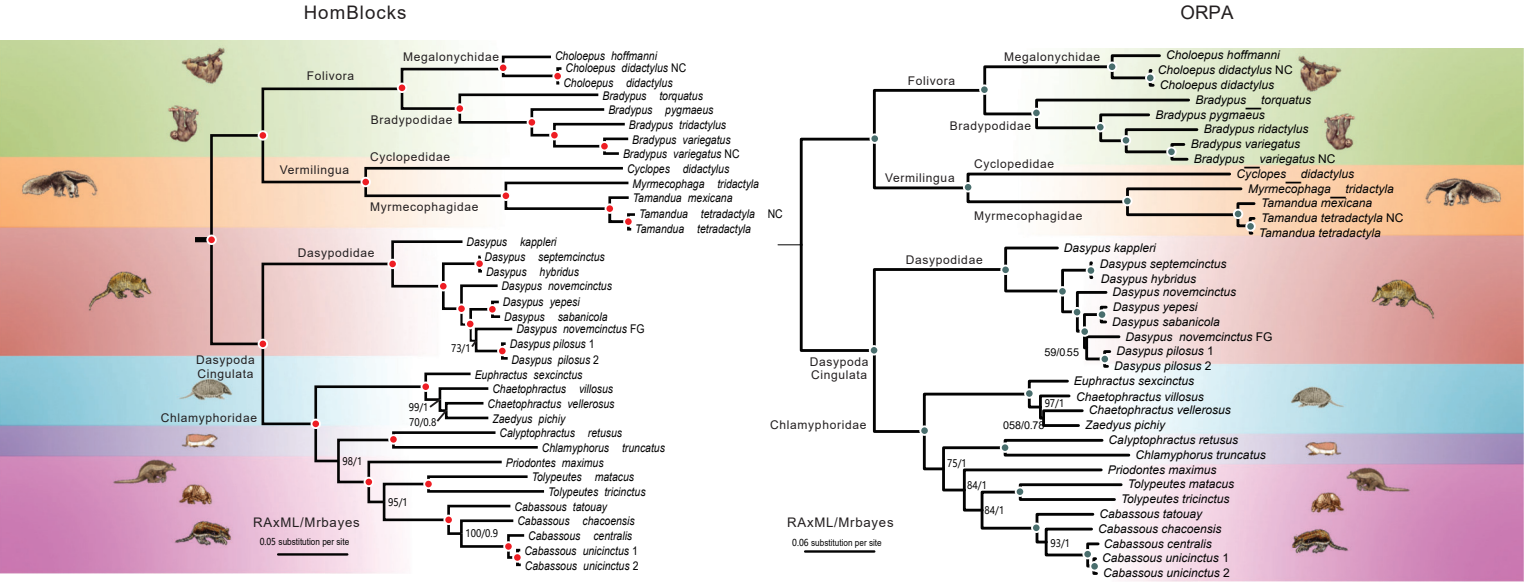
Fig. 4 Comparison of phylogenetic trees and alignment methods for 18 higher plant mitochondrial genomes. a, Two phylogenetic trees of 18 higher plant mitochondrial genomes were constructed using ORPA and HomBlocks alignment methods, respectively. The trees were

277 constructed with maximum likelihood and Bayesian inference methods, and the support values
278 derived from RAxML and Bayesian posterior probability are indicated on each node. Fully
279 resolved nodes are unlabeled. **b**, Distributional differences of phylogenetic alignments obtained
280 from ORPA and HomBlocks methods using Ajuga reptans as the reference sequence. The circos
281 plot illustrates the differing sequence composition sites between the two methods, with green and
282 gray dots indicating the variation between the alignments.
283
284 **Fig. 5 Comparison of ORPA and HomBlocks runtime efficiency**. **a**, Comparison of runtime
285 for 60 higher plant chloroplast datasets. A maximum likelihood tree shows the evolutionary
286 relationship among 60 samples. Nodes with 100% support are unspecified, and other partially
287 supported nodes are labeled with bootstrap and aLTR values. Sampling begins at the base of the
288 tree and proceeds with increasing sample sizes of 5 until all data are used, resulting in a total of
289 12 comparison groups. **b**, Comparison of ORPA and HomBlocks runtime. The sample size
290 corresponds to the sampling range in Figure 5a. The percentage on the bar chart represents the
291 similarity in systemic tree topology generated by the two software programs.
292
293 **Fig. 6 Identification of species-level evolutionary conflicts using ORPA.** A total of 52
294 Lamiales chloroplast trees were constructed using 101,544 characters from the ORPA alignment.
295 Maximum likelihood and Bayesian inference methods were used to construct the trees, and the
296 support values derived from RAxML (left) and Bayesian posterior probability (right) are
297 indicated by numerical values on the nodes. Fully resolved nodes are indicated by red dots.
298 *Wightia speciosissima*, which has a controversial position in Lamiales, is labelled in red. The
299 morphology of four species from *Schrebera* and *Comoranthus* genera is shown on the right side

300 of the figure. Additionally, the results reveal a paraphyletic relationship, with *Comoranthus*
301 *minor* nested within *Schrebera*, leading to the synonymization of these genera.

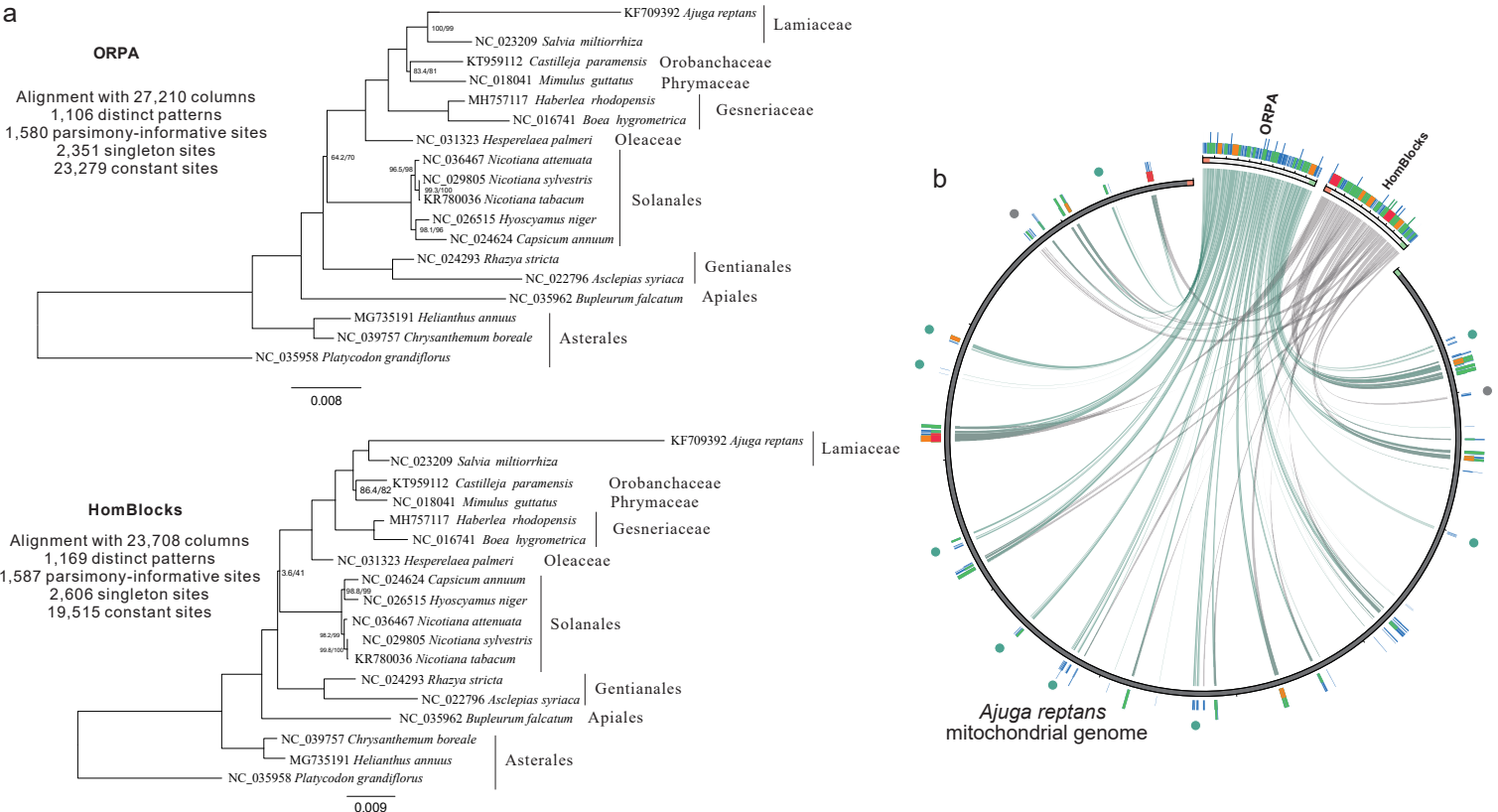

HomBlocks: 0.5~30 hours
Sequencing
Data collection (Sampling)
Gene alignment
Super matrix
Tree building
Organelle Genomes
A solid tree
Sequencing
Data collection (Sampling)
Gene alignment
Super matrix
Tree building
ORPA: 5 minutes

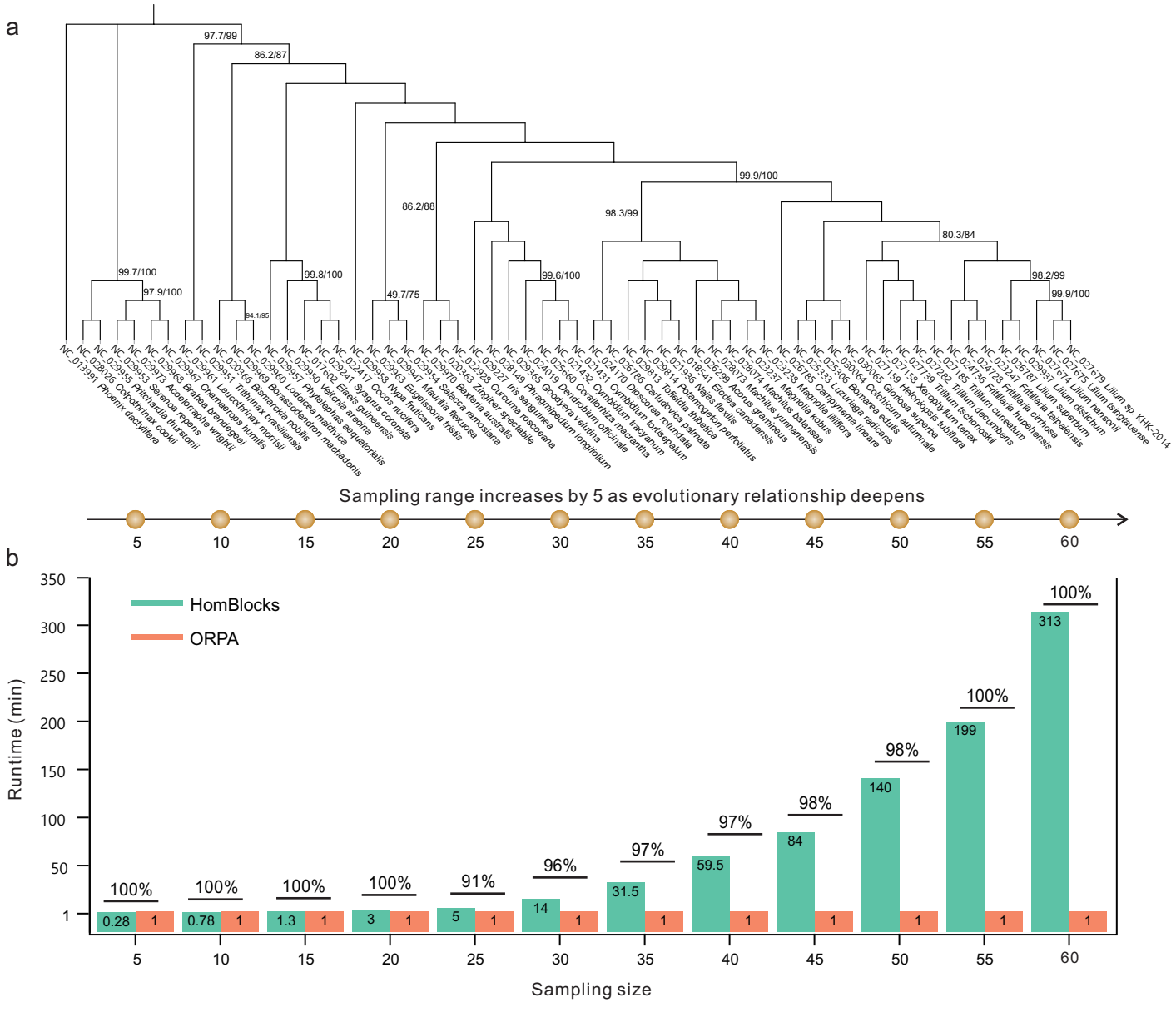

HomBlocks
Alignment with 62,101 columns
5,824 distinct patterns
8,404 parsimony-informative sites
7,378 singleton sites
46,319 constant sites
ORPA
Alignment with 90,925 columns
5,459 distinct patterns
8,270 parsimony-informative sites
8,679 singleton sites
73,976 constant sites
Aegilops cylindrica
Aegilops tauschii
Aegilops geniculata
Triticum monococcum
Triticum urartu
Triticum aestivum
Aegilops speltoides cv.SPE0661
Secale cereale
Hordeum vulgare ssp. vulgare
Lolium perenne
Lolium multiflorum
Festuca pratensis
Festuca arundinacea
Festuca altissima
Festuca ovina
Deschampsia antarctica
Agrostis stolonifera
Brachypodium distachyon
Aegilops speltoides
Hordeum vulgare
Pooideae
Phyllostachys edulis
Phyllostachys nigra var. henonii
Phyllostachys propinqua
Indocalamus longiauritus
Arundinaria tecta
Arundinaria appalachiana
Arundinaria gigantea
Acidosasa purpurea
Ferrocalamus rimosivaginus
Bambusa oldhamii
Bambusa multiplex
Bambusa emeiensis
Dendrocalamus latiflorus
45/1
45/0.5806
98/1
72/1
82/1
82/1
Bambusoideae
Oryza sativa cv. 93-11
Oryza nivara
Oryza rufipogon
Oryza sativa cv. Nipponbare
Oryza meridionalis
Leersia tisserantii
Rhynchoryza subulata
Zizania latifolia
Oryza sativa
88/1
97/1
63/1
Ehrartoideae
Sorghum bicolor
Sorghum timorense
Saccharum officinarum cv. NCo310
Coix lacryma-jobi
Zea mays
Panicum virgatum
Setaria italica
Phragmites australis
Panicoideae
Node supports
RAxML/Mrbayes
Puelia olyriformis
Pharus lappulaceus
Pharus latifolius
Arundinoideae
Puelioideae
Pharoideae
Anomochlooideae
Anomochloa marantoidea
Typha latifolia
0.01
0.01

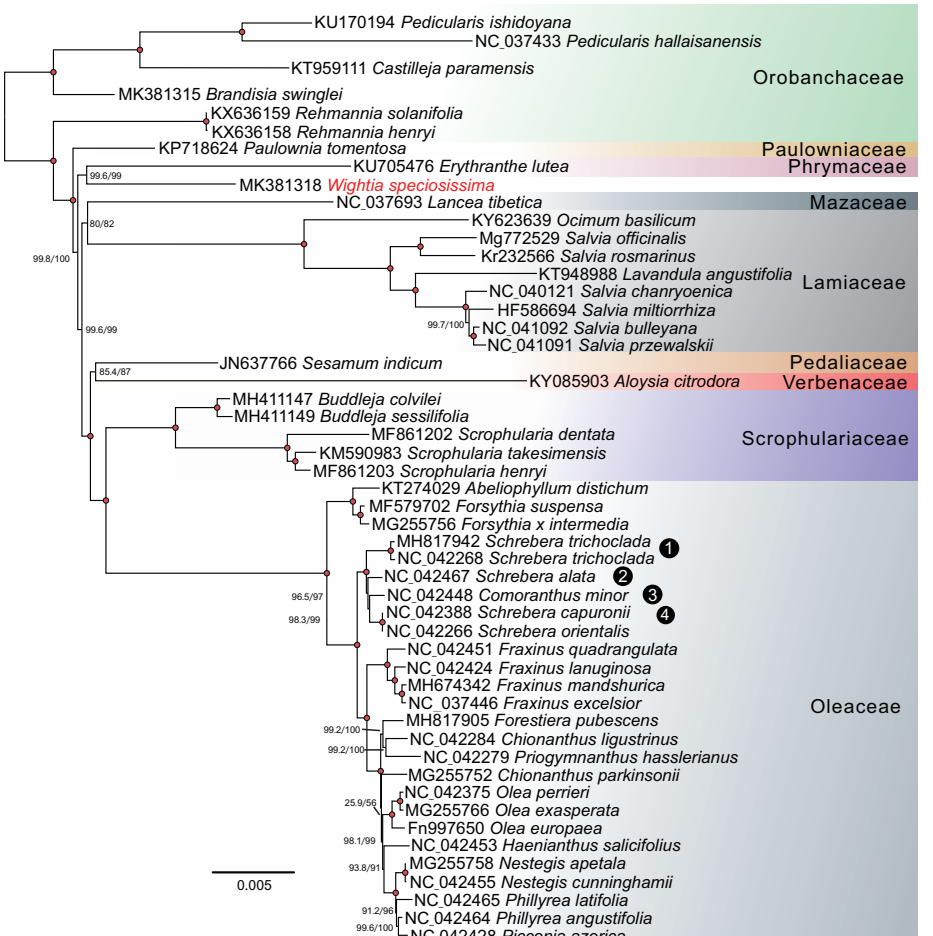


Alignment has 15,170 columns, 6,201 distinct patterns
6,722 parsimony-informative sites, 919 singleton sites, 7,529 constant sites

Alignment 8,696 columns, 2,891 distinct patterns
3,015 parsimony-informative sites, 494 singleton sites, 5,187 constant sites


a
ORPA
Alignment with 27,210 columns
1,106 distinct patterns
1,580 parsimony-informative sites
2,351 singleton sites
23,279 constant sites
KF709392 Ajuga reptans
NC_023209 Salvia miltiorrhiza
KT959112 Castilleja paramensis
NC_018041 Mimulus guttatus
MH757117 Haberlea rhodopensis
NC_016741 Boea hygrometrica
NC_031323 Hesperelaea palmeri
NC_036467 Nicotiana attenuata
NC_029805 Nicotiana sylvestris
KR780036 Nicotiana tabacum
NC_026515 Hyoscyamus niger
NC_024624 Capsicum annuum
NC_024293 Rhazya stricta
NC_022796 Asclepias syriaca
NC_035962 Bupleurum falcatum
MG735191 Helianthus annuus
NC_039757 Chrysanthemum boreale
NC_035958 Platycodon grandiflorus
Lamiaceae
Orobanchaceae
Phrymaceae
Gesneriaceae
Oleaceae
Solanales
Gentianales
Apiales
Asterales
100/99
83.4/81
64.2/70
96.5/98
99.3/100
98.1/96
0.008
HomBlocks
Alignment with 23,708 columns
1,169 distinct patterns
1,587 parsimony-informative sites
2,606 singleton sites
19,515 constant sites
KF709392 Ajuga reptans
NC_023209 Salvia miltiorrhiza
KT959112 Castilleja paramensis
NC_018041 Mimulus guttatus
MH757117 Haberlea rhodopensis
NC_016741 Boea hygrometrica
NC_031323 Hesperelaea palmeri
NC_024624 Capsicum annuum
NC_026515 Hyoscyamus niger
NC_036467 Nicotiana attenuata
NC_029805 Nicotiana sylvestris
KR780036 Nicotiana tabacum
NC_024293 Rhazya stricta
NC_022796 Asclepias syriaca
NC_035962 Bupleurum falcatum
NC_039757 Chrysanthemum boreale
MG735191 Helianthus annuus
NC_035958 Platycodon grandiflorus
Lamiaceae
Orobanchaceae
Phrymaceae
Gesneriaceae
Oleaceae
Solanales
Gentianales
Apiales
Asterales
86.4/82
3.6/41
98.8/99
98.2/99
99.8/100
0.009
b
ORPA
HomBlocks
Ajuga reptans
mitochondrial genome


a
97.7/99
86.2/87
99.9/100
86.2/88
98.3/99
80.3/84
99.7/100
97.9/100
99.8/100
49.7/75
99.6/100
98.2/99
99.9/100
94.1/95
Nc_013991 Phoenix dactylifera
Nc_030036 Calophrynium cookii
Nc_029955 Phrynium pubinerve
Nc_029953 Sarcenca repens
Nc_029972 Acoelorraphe wrightii
Nc_029968 Brahea brandegeei
Nc_029967 Chamaerops humilis
Nc_029961 Leucothrinax morrisii
Nc_029951 Trithrinax brasiliensis
Nc_029966 Bismarckia nobilis
Nc_029969 Borassodendron machadonis
Nc_029960 Lodoicea maldivica
Nc_029957 Phytelephas aequatorialis
Nc_017602 Veitchia arecina
Nc_029241 Elaeis guineensis
Nc_022417 Syagrus coronata
Nc_029958 Nypa fruticans
Nc_029963 Cocos nucifera
Nc_029947 Eugeissona tristis
Nc_029954 Maurita flexuosa
Nc_029970 Salacca amesiana
Nc_030363 Bactris australis
Nc_029227 Zingiber spectabile
Nc_029228 Curcuma roscoeana
Nc_029149 Iris sanguinea
Nc_030365 Phragmipedium longifolium
Nc_024019 Goodyera velutina
Nc_025660 Dendrobium officinale
Nc_031432 Coralorrhiza maculata
Nc_031431 Cymbidium macrantha
Nc_024170 Cymbidium tracyanum
Nc_026786 Dioscorea rotundata
Nc_029813 Carludovica palmata
Nc_029814 Toliellia palmata
Nc_031936 Nagas flexilis
Nc_018541 Elodea canadensis
Nc_029299 Aconis gramineus
Nc_030073 Machilus yunnanensis
Nc_030074 Machilus thunbergii
Nc_022237 Magnolia kobus
Nc_029238 Magnolia biondii
Nc_026185 Magnolia sprengeri
Nc_033333 Liriodendron chinense
Nc_025306 Liquidambar formosana
Nc_030064 Bomarea edulis
Nc_030065 Calycanthus floridus
Nc_027159 Heloniopsis orientalis
Nc_027158 Xerophyllum tenax
Nc_027739 Trillium tschonoskii
Nc_027282 Trillium camschatcense
Nc_027185 Trillium decumbens
Nc_024736 Trillium cuneatum
Nc_024728 Fritillaria imperialis
Nc_032401 Fritillaria cirrhosa
Nc_026187 Fritillaria unibracteata
Nc_029931 Lilium superbum
Nc_027674 Lilium lankongense
Nc_027675 Lilium hansonii
Nc_027679 Lilium dauricum
Nc_027680 Lilium sp. KHK-2014
Sampling range increases by 5 as evolutionary relationship deepens
5
10
15
20
25
30
35
40
45
50
55
60
b
HomBlocks
ORPA
Runtime (min)
Sampling size
350
300
250
200
150
100
50
1
100%
100%
100%
100%
91%
96%
97%
97%
98%
98%
100%
100%
0.28
0.78
1.3
3
5
14
31.5
59.5
84
140
199
313
1



Schrebera trichoclada

Schrebera alata

Comoranthus minor

Schrebera capuronii