**The first fully-annotated chloroplast sequence of the large pantropical family Annonaceae**

**Running title: The first Annonaceae plastome sequence**

Authors: Rita D. Brandão1,\*, Martijn Staats2, Lars W. Chatrou2, Jan W. Maas3, Hans Persoon4, Tatiana Arias5, Roy H.J. Erkens1,6

1. Maastricht Science Programme, Maastricht University, The Netherlands
2. Biosystematics group, Wageningen University, Wageningen, The Netherlands
3. Ecology and Biodiversity Group, Department of Biology, Utrecht University, Utrecht, The Netherlands
4. Botanical Garden Utrecht, Utrecht University, Utrecht, The Netherlands
5. School of Biological Sciences, The University of Hong Kong, China
6. Naturalis Biodiversity Center, Leiden, The Netherlands

\*Corresponding author:

rita.brandao@maastrichtuniversity.nl; PO Box 616, 6200 MD Maastricht; tel. +31 043 3881350

**Abstract**

The pantropical flowering plant family Annonaceae is one of the largest families within the Magnoliidae with c. 2,400 species. The study of its evolution allows gaining insight into the evolution of basal angiosperms, the evolution of rain forests and biodiversification. Phylogenetic analyses have greatly contributed to these studies. However, these are so far limited to few DNA markers. Here we present the fully-annotated chloroplast genome (plastome) sequence from an Annonaceae species: *Uvaria afzelii* and we compare its organizational features with those from other nine Annonaceae species and 8 basal angiosperm species.

The 167,530 bp plastome of *Uvaria afzelii* contains 139 genes and an inversion within the LSC (57,722 - 66,951 bp) compared to the chloroplasts of *Magnolia kwansiensis* and *Liriodendron tulipifera.* In general, the organization seems to be similar to that of other basal angiosperms and Annonaceae, including our *Annona cherimola* sample, whereas it is strikingly different from the previously reported *Annona cherimola*. The differences found for the organization of the plastome between the two *Annona cherimola* could derive from the fact that they are from different cultivars.

Analysis of the Annonaceae plastomes will reveal useful information for improved phylogeographic analyses, species-level DNA barcoding and general understanding of magnoliid evolution as well as rain forest evolution.

Keywords: Annonaceae; next-generation sequencing; plastome sequence; complete chloroplast sequence; phylogenetics

Understanding the origin of flowering-plant biodiversity is one of the major biological research themes of the moment. A lot of effort has been devoted to studying the monocot and eudicot lineages but less to the other large clade of flowering plants, Magnoliidae, although it contains c. 4% of all plant species. Within Magnoliidae, the pantropically distributed Annonaceae is a genus-rich family, with 109 genera and *c*. 2,400 accepted species ([Chatrou et al., 2012](#_ENREF_7)). Understanding the evolution and diversification of Annonaceae will therefore greatly aid in understanding the evolution of basal angiosperms. Additionally, species richness and abundance of individuals within the tropical rainforests indicate Annonaceae may be a key group to understand the evolution of tropical diversity ([van Gemerden et al., 2003](#_ENREF_37), [Gentry, 1988](#_ENREF_13), [Valencia et al., 1994](#_ENREF_36)).

Molecular phylogenetic studies in Annonaceae, as in many other clades, have mostly been conducted using few chloroplast markers whether it be at the family, subfamily, tribal ([Pirie et al., 2006](#_ENREF_24), [Erkens et al., 2009](#_ENREF_12), [Chaowasku et al., 2014](#_ENREF_5)), or generic levels ([Su and Saunders, 2009](#_ENREF_33), [Erkens et al., 2014](#_ENREF_10), [Chatrou et al., 2012](#_ENREF_7)), with few exceptions ([Chatrou et al., 2009](#_ENREF_6), [Thongpairoj, 2008](#_ENREF_35)). The suite of plastid markers available for Annonaceae is limited, mainly due to the fact that previous knowledge of the chloroplast DNA sequence is needed in order to identify the most informative ones at different taxonomic levels. Nowadays, the advent of next generation sequencing (NGS) techniques is revolutionizing the field of phylogenetics. NGS allows us to generate massive amounts of DNA sequencing data in order to revisit challenging evolutionary questions that could not be answered in a time efficient manner using Sanger sequencing ([Steele et al., 2012](#_ENREF_31), [Straub et al., 2012](#_ENREF_32)). So, it is possible to sequence entire plastomes of these plants, including from herbarium material using methods such as published by Staats et al.([Staats et al., 2013](#_ENREF_30)) Whole chloroplast assembly using shallow genomic DNA sequencing has become a popular method among botanists for reconstructing plant phylogenies ([Steele et al., 2012](#_ENREF_31)). NGS sequencing is cheaper and faster than traditional PCR-based methods to obtain large amounts of data. These large amounts of data can also be used to identify more relevant regions for resolving phylogenies of difficult clades ([Bortiri et al., 2008](#_ENREF_2), [Huang et al., 2014](#_ENREF_14), [Matsuoka et al., 2002](#_ENREF_22), [Shaw et al., 2007](#_ENREF_28)). Currently, in the RefSeq Database ([Pruitt et al., 2002 [updated 2012]](#_ENREF_26)) there are almost 900 plant chloroplast genome entries, of which 685 belong to flowering plants ([Lange et al., 2008](#_ENREF_20))([Lange et al., 2008](#_ENREF_20))(Genbank, accessed May 2016). Basal angiosperms (ANITA clade, Magnoliidae clade, Choranthaceae and Ceratophyllaceae) are only represented by 29 plastomes, which represents 4% of all angiosperms' plastomes. Despite the increasing number of chloroplast genomes every year, until this study, no chloroplast genome was available for Annonaceae. Here, we describe the first plastome of a species of Annonaceae: *Uvaria afzelii* G.Elliott (tribe Uvarieae, subfamily Annonoideae ([Chatrou et al., 2012](#_ENREF_7))). Small trees or spreading shrubs of *Uvaria afzelii* may grow up to 5 metres tall. This species is distributed from Guinea to southern Nigeria; the fruit is edible and leaves, bark and roots are used for their medicinal properties ([Burkill, 1985](#_ENREF_3)). The chloroplast genomes of *several* Annonaceae will aid to the targeted-sequencing approach of other Annonaceae species and mapping since these sequences can be used as closely related reference genomes for assembly. Furthermore, it allows identification of new markers that may help resolving difficult clades, or DNA targeting enrichment for sequencing approaches.

**Materials and methods**

DNA extraction and library preparation

Fresh leaf material from *Uvaria afzelii* was collected from a green house grown tree at the Botanical garden of Utrecht University (Utrecht, The Netherlands). The chloroplast isolation kit (Sigma, Saint-Louis, USA) was used according to the manufacturer’s instructions. DNA extraction from the chloroplasts was performed with the GenElute plant genomic DNA miniprep kit (Sigma). The obtained DNA was sent to Macrogen (Korea) for paired-end library construction and sequencing on a HiSeq 2000 (Illumina) instrument, following the company’s protocol.

Sequencing analysis

Quality control was performed with FastQC and FastQ Screen (both available at http://www.bioinformatics.babraham.ac.uk/projects/download.html). Geneious version 6.06 ([Kearse et al., 2012](#_ENREF_17)) ([http://www.geneious.com](http://www.geneious.com/)) was used to filter poor quality bases and/or reads and iterative mapping steps. In the first mapping step, reads were mapped to the draft sequence of a chloroplast of *Miliusa cuneata* Craib, an Asian Annonaceae species (kindly provided by Arias et al., unpublished data), to assemble a draft of the *U. afzelii* plastome. In the following 20 steps, consecutive rounds of mapping to the obtained sequences were performed. Remaining gaps were resolved in two ways: a) *in silico* primer walking using the *de novo* assembler PRICE TI ([Ruby et al., 2013](#_ENREF_27)) on previously filtered data from the software TrimGalore! (available at http://www.bioinformatics.babraham.ac.uk/projects/download.html); b) PCRs using primers flanking the gaps followed by automated Sanger sequencing.

Gene annotation and chloroplast circular map design

For Uvaria afzelli, gene annotation was performed in Geneious software upon alignment of the complete chloroplast sequences of *Liriodendron tulipifera* ([Cai et al., 2006](#_ENREF_4)) (genbank accession number: NC\_008326.1) and *Magnolia kwangsiensis* ([Kuang et al., 2011](#_ENREF_19)) (genbank accession number: NC\_015892.1) using the MAFF package, and in DOGMA ([Wyman et al., 2004](#_ENREF_39)). The exact boundaries between the inverted repeats and the single copies were determined manually. GenomeVx ([Conant and Wolfe, 2008](#_ENREF_8)) was used to draw the circular map of the chloroplast genome. The complete chloroplast genome with annotations was submitted to Genbank (accession number XXXXXXX). Other Annonaceae plastomes were annonated with Geneious, using Uvaria afzelii as reference.

**Results and discussion**

Results of the quality control freeware Fastq Screen on the raw sequence data show that only a small percentage of all 20,636,570 reads belong to the chloroplast sequence (figure 1). Less than 0.6% of the total number of reads was aligned to an unpublished draft plastome of *Miliusa cuneata* (Arias et al., unpublished data), which was used as a reference. The majority of the reads did not correspond to human, mouse or insect DNA, therefore are inferred to be nuclear and mitochondrial plant DNA. After assembly of the reads by mapping to the reference chloroplast, several iteration steps were performed, increasing the percentage of used reads to 0.7%. Few gaps with coverage below 2X were identified and resolved using both *in silico* primer walking and PCR followed by automated Sanger sequencing method.

The complete chloroplast genome of *Uvaria afzelii* (figure 2) is around 167,530 bp. It is noteworthy that small indels, as well as SNPs, were observed in heteroplasmy, so the actual genome size varies slightly. The inverted repeats (IRs) have 28,002 and 28.095 bp respectively and are separated by the small single-copy region (20,010 bp) and the long single-copy region (91,296 bp). We found 139 genes in the chloroplast DNA of *Uvaria afzelii*, 21 of which are completely duplicated and 2 partially duplicated within the IRs. The most noteworthy features of its chloroplast genome is an inversion within the LSC (57,722 - 66,951 bp) compared to the chloroplasts of *Magnolia kwansiensis* ([Kuang et al., 2011](#_ENREF_19))and *Liriodendron tulipifera* ([Cai et al., 2006](#_ENREF_4)). This region includes the following genes: atpE, atpB, rbcL, accD, psaI and ycf4.

We then obtained and annotated additional Annonaceae genomes: *Anaxagorea phaeocarpa Mart vel aff*, *Annona muricata*, *Annona cherimola*, *Asimina pygmaea*, *Cananga kirkii*, *Crematosperma leiophyllum*, *Oxandra macrophylla*, and *Xylopia peruviana*. We have compared the plastomes’ structures with plastomes from ancestral Angiosperms and the recently published plastome of *Anonna cherimola*. The results are shown in Table 1. Interestingly, the inversion described above that spans over 6 genes seems to be Annonaceae specific.

CpDNA heteroplasmy, the condition in which there is more than one organellar haplotype, has been described in several angiosperms genera, such as: *Medicago* ([Johnson and Palmer](#_ENREF_15)) (Fabaceae), *Coreopsis* ([Mason et al., 1994](#_ENREF_21))(Asteraceae), *Turnera* ([Shore et al., 1994](#_ENREF_29)) (Passifloraceae). This event poses challenges for phylogenetic analyses as bioinformatic tools are not ready to deal with this event. The use of NGS to sequence (parts of) genomes at high coverage allows easily detection of these regions. So, polymorphism-aware phylogenetic tools must be implemented to account for heteroplasmy.

It is important to realise that despite the chloroplast circular structure representation and the fact that Genbank accessions mention that the chloroplast genome is circular, the chloroplast DNA mostly has a complex and branched form, and the circular form is only present in small amounts ([Bendich, 2004](#_ENREF_1)). Probably due to this particularity of the plastome molecule, PCRs over the IRb/LSC boundary, where the chloroplast should circularize, were not successful.

Chloroplast DNA is a very gene-rich molecule, with more than 100 genes, spanning over 100-200 kb comparing plant mitochondrial genome (around 60 genes in a variable size with a minimum of 200kb). Chloroplast genes code for the entire machinery necessary for the photosynthetic process. There is a relatively high degree of conservation in size, structure, gene content, and linear order of the chloroplast genes in land plants ([Downie and Palmer, 1992](#_ENREF_9)). In chloroplasts, DNA replication is cell-cycle independent and the number of chloroplast copies present per cell varies. In rapidly dividing leaf tissue there are around hundred copies, but later in leaf development and plant growth, the number of copies decreases to about 20-30 ([Possingham, 1980](#_ENREF_25)). In addition, each chloroplast may contain up to 300 copies of the genome ([Krupinska et al., 2013](#_ENREF_18)). Despite the abundance of the chloroplast genome, genome skimming (shallow sequencing) experiments show that only a small percentage of the reads actually map to the chloroplast. So, targeting enrichment sequencing is a very promising technique to sequence the plastome from many different species. However, unlike the standard NGS sequencing protocols, this technique requires some *a priori* knowledge of the genome sequence, such as the genome of closely related species. The sequencing of Annonaceae chloroplast genomes, such as the one presented here will now allow for the construction of (homemade) plastome-target probes for the sequencing of many chloroplast genomes. Parks et al. showed that nearly complete plastomes of 37 *Pinus* species allowed for an increased phylogenetic resolution and support values at lower taxonomic levels ([Parks et al., 2009](#_ENREF_23)). Similar results were obtained by others using plastome sequences alone ([Whittall et al., 2010](#_ENREF_38), [Yang et al., 2013](#_ENREF_40)) or in combination with nuclear ribosomal sequences ([Kane et al., 2012](#_ENREF_16)). Therefore, complete Annonaceae plastome analysis might also reveal useful for improved phylogeographic analyses to elucidate relationships within, for instance, recently radiated genera ([Erkens et al., 2007](#_ENREF_11), [Tang et al., 2015](#_ENREF_34)) or ill-resolved branching events in the Annonaceae family phylogeny ([Erkens et al., 2009](#_ENREF_12)). Lastly, species-level DNA barcoding and general understanding of magnoliid evolution as well as investigations of rain forest evolution greatly benefit from this data.

**Acknowledgements:** This work was supported in part by an Innovational Research Incentives Scheme grant (VENI, nr. 863.09.017; NWO-ALW) to RHJE.

**References**

Bendich AJ. 2004. Circular Chloroplast Chromosomes: The Grand Illusion. *The Plant Cell Online* 16: 1661-1666.

Bortiri E, Coleman-Derr D, Lazo GR, Anderson OD,Gu YQ. 2008. The complete chloroplast genome sequence of *Brachypodium distachyon*: sequence comparison and phylogenetic analysis of eight grass plastomes. *BMC Research Notes* 1: 61-61.

Burkill HM 1985. *The useful plants of West Tropical Africa,* Kew, Richmond, United Kingdom, Royal Botanic Gardens.

Cai Z, Penaflor C, Kuehl J, Leebens-Mack J, Carlson J, dePamphilis C, Boore J,Jansen R. 2006. Complete plastid genome sequences of *Drimys*, *Liriodendron*, and *Piper*: implications for the phylogenetic relationships of magnoliids. *BMC Evolutionary Biology* 6: 77.

Chaowasku T, Thomas DC, van der Ham RWJM, Smets EF, Mols JB,Chatrou LW. 2014. A plastid DNA phylogeny of tribe Miliuseae: Insights into relationships and character evolution in one of the most recalcitrant major clades of Annonaceae. *American Journal of Botany* 101: 691-709.

Chatrou LW, Escribano MP, Viruel MA, Maas JW, Richardson JE,Hormaza JI. 2009. Flanking regions of monomorphic microsatellite loci provide a new source of data for plant species-level phylogenetics. *Molecular Phylogenetics and Evolution* 53: 726-733.

Chatrou LW, Pirie MD, Erkens RHJ, Couvreur TLP, Neubig KM, Abbott JR, Mols JB, Maas JW, Saunders RMK,Chase MW. 2012. A new subfamilial and tribal classification of the pantropical flowering plant family Annonaceae informed by molecular phylogenetics. *Botanical Journal of the Linnean Society* 169: 5-40.

Conant GC,Wolfe KH. 2008. GenomeVx: simple web-based creation of editable circular chromosome maps. *Bioinformatics* 24: 861-862.

Downie S,Palmer J 1992. Use of chloroplast DNA rearrangements in reconstructing plant phylogeny. *In:* SOLTIS, P., SOLTIS, D. & DOYLE, J. (eds.) *Molecular Systematics of Plants.* Springer US.

Erkens RHJ, Chatrou LW, Chaowasku T, Westra LYT, Maas JW,Maas PJM. 2014. A decade of uncertainty: Resolving the phylogenetic position of *Diclinanona* (Annonaceae), including taxonomic notes and a key to the species. *Taxon* 63: 1244–1252.

Erkens RHJ, Chatrou LW, Maas JW, van der Niet T,Savolainen V. 2007. A rapid diversification of rainforest trees (*Guatteria*; Annonaceae) following dispersal from Central into South America. *Molecular Phylogenetics and Evolution* 44: 399-411.

Erkens RHJ, Maas JW,Couvreur TLP. 2009. From Africa via Europe to South America: migrational route of a species-rich genus of Neotropical lowland rain forest trees (*Guatteria*, Annonaceae). *Journal of Biogeography* 36: 2338-2352.

Gentry AH. 1988. Tree species richness of upper Amazonian forests. *Proceedings of the National Academy of Sciences* 85: 156-159.

Huang H, Shi C, Liu Y, Mao S-Y,Gao L-Z. 2014. Thirteen *Camellia* chloroplast genome sequences determined by high-throughput sequencing: genome structure and phylogenetic relationships. *BMC Evolutionary Biology* 14: 151.

Johnson LB,Palmer JD. Heteroplasmy of chloroplast DNA in *Medicago*. *Plant Molecular Biology* 12: 3-11.

Kane N, Sveinsson S, Dempewolf H, Yang JY, Zhang D, Engels JMM,Cronk Q. 2012. Ultra-barcoding in cacao (Theobroma spp.; Malvaceae) using whole chloroplast genomes and nuclear ribosomal DNA. *American Journal of Botany* 99: 320-329.

Kearse M, Moir R, Wilson A, Stones-Havas S, Cheung M, Sturrock S, Buxton S, Cooper A, Markowitz S, Duran C, Thierer T, Ashton B, Meintjes P,Drummond A. 2012. Geneious Basic: An integrated and extendable desktop software platform for the organization and analysis of sequence data. *Bioinformatics* 28: 1647-1649.

Krupinska K, Melonek J,Krause K. 2013. New insights into plastid nucleoid structure and functionality. *Planta* 237: 653-664.

Kuang D-Y, Wu H, Wang Y-L, Gao L-M, Zhang S-Z,Lu L. 2011. Complete chloroplast genome sequence of *Magnolia kwangsiensis* (Magnoliaceae): implication for DNA barcoding and population genetics. *Genome* 54: 663-673.

Lange EM, Sun J, Lange LA, Zheng SL, Duggan D, Carpten JD, Gronberg H, Isaacs WB, Xu J,Chang B-L. 2008. Family-Based Samples Can Play an Important Role in Genetic Association Studies. *Cancer Epidemiol Biomarkers Prev* 17: 2208-2214.

Mason RJ, Holsinger KE,Jansen RK. 1994. Biparental inheritance of the chloroplast genome in *Coreopsis* (Asteraceae). *Journal of Heredity* 85: 171-173.

Matsuoka Y, Yamazaki Y, Ogihara Y,Tsunewaki K. 2002. Whole Chloroplast Genome Comparison of Rice, Maize, and Wheat: Implications for Chloroplast Gene Diversification and Phylogeny of Cereals. *Molecular Biology and Evolution* 19: 2084-2091.

Parks M, Cronn R,Liston A. 2009. Increasing phylogenetic resolution at low taxonomic levels using massively parallel sequencing of chloroplast genomes. *BMC Biology* 7: 1-17.

Pirie MD, Chatrou LW, Mols JB, Erkens RHJ,Oosterhof J. 2006. ‘Andean-centred’ genera in the short-branch clade of Annonaceae: testing biogeographical hypotheses using phylogeny reconstruction and molecular dating. *Journal of Biogeography* 33: 31-46.

Possingham JV. 1980. Plastid replication and development in the life cycle of higher plants. *Annual Review of Plant Physiology* 31: 113-129.

Pruitt K, Brown G, Tatusova T,Maglott D 2002 [updated 2012]. The Reference Sequence (RefSeq) database. *The NCBI Handbook [Internet].* Bethesda (MD): National Center for Biotechnology Information (US).

Ruby JG, Bellare P,DeRisi JL. 2013. PRICE: Software for the targeted assembly of components of (meta) genomic sequence data. *G3: Genes|Genomes|Genetics* 3: 865-880.

Shaw J, Lickey EB, Schilling EE,Small RL. 2007. Comparison of whole chloroplast genome sequences to choose noncoding regions for phylogenetic studies in angiosperms: the tortoise and the hare III. *American Journal of Botany* 94: 275-288.

Shore JS, McQueen KL,Little SH. 1994. Inheritance of plastid DNA in the *Turnera ulmifolia* complex (Turneraceae). *American Journal of Botany* 81: 1636-1639.

Staats M, Erkens RHJ, van de Vossenberg B, Wieringa JJ, Kraaijeveld K, Stielow B, Geml J, Richardson JE,Bakker FT. 2013. Genomic treasure troves: complete genome sequencing of herbarium and insect museum specimens. *PLoS ONE* 8: e69189.

Steele PR, Hertweck KL, Mayfield D, McKain MR, Leebens-Mack J,Pires JC. 2012. Quality and quantity of data recovered from massively parallel sequencing: Examples in Asparagales and Poaceae. *American Journal of Botany* 99: 330-348.

Straub SCK, Parks M, Weitemier K, Fishbein M, Cronn RC,Liston A. 2012. Navigating the tip of the genomic iceberg: Next-generation sequencing for plant systematics. *American Journal of Botany* 99: 349-364.

Su Y,Saunders R. 2009. Evolutionary divergence times in the Annonaceae: evidence of a late Miocene origin of *Pseuduvaria* in Sundaland with subsequent diversification in New Guinea. *BMC Evolutionary Biology* 9: 153.

Tang CC, Thomas DC,Saunders RMK. 2015. Molecular phylogenetics of the species-rich angiosperm genus *Goniothalamus* (Annonaceae) inferred from nine chloroplast DNA regions: Synapomorphies and putative correlated evolutionary changes in fruit and seed morphology. *Molecular Phylogenetics and Evolution* 92: 124-139.

Thongpairoj U-S. 2008. *Taxonomy and molecular phylogeny of Artabotrys R.Brown and palynology of the tribe Unoneae (Annonaceae).* PhD PhD thesis, Chiang Mai University, Thailand.

Valencia R, Balslev H,Paz Y Miño C G. 1994. High tree alpha-diversity in Amazonian Ecuador. *Biodiversity & Conservation* 3: 21-28.

van Gemerden BS, Olff H, Parren MPE,Bongers F. 2003. The pristine rain forest? Remnants of historical human impacts on current tree species composition and diversity. *Journal of Biogeography* 30: 1381-1390.

Whittall JB, Syring J, Parks M, Buenrostro J, Dick C, Liston A,Cronn R. 2010. Finding a (pine) needle in a haystack: chloroplast genome sequence divergence in rare and widespread pines. *Molecular Ecology* 19: 100-114.

Wyman SK, Jansen RK,Boore JL. 2004. Automatic annotation of organellar genomes with DOGMA. *Bioinformatics* 20: 3252-3255.

Yang J-B, Tang M, Li H-T, Zhang Z-R,Li D-Z. 2013. Complete chloroplast genome of the genus *Cymbidium*: lights into the species identification, phylogenetic implications and population genetic analyses. *BMC Evolutionary Biology* 13: 1-12.

|  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
|  | **Table 1 – Comparison of plastome sizes and organization characteristics** | | | | | | | | | | | | | | | |
| **Species name** | | **Genbank**  **acession number** | **Total genome** | | | | | | | | **LSC** | | **SSC** | | **IR** | |
| **Size**  **(bp)** | **# genes** | **+ Number** | **- Number** | **#Protein-coding genes** | **+ Number** | **- Number** | **Inversion\*** | **Size**  **(bp)** | **# genes** | **Size**  **(bp)** | **# genes** | **Size**  **(bp)** | **# genes** |
| *Amborella trichopoda* | | NC\_005086 | 162 686 | 128 | 50 | 78 | 84 | 29 | 55 | No | 90 970 | 85 | 18 415 | 13 | 26 651 | 16 |
| *Cerathophyllum demersum* | | NC\_009962 | 156 252 | 130 | 50 | 80 | 85 | 48 | 55 | No | 86 607 | 85 | 19 007 | 13 | 25 319 | 16 |
| *Chloranthus japonicus* | | NC\_026565 | 158 640 | 131 | 57 | 74 | 86 | 59 | 54 | No | 87 724 | 84 | 18 618 | 13 | 26 149 | 17 |
| *Illicium oligandrum* | | NC\_009600 | 148 553 | 126 | 47 | 79 | 83 | 45 | 35 | No | 98 057 | 90 | 20 267 | 13 | 15 115 | 12 |
| *Liriodrendon tulipifera* | | NC\_008326 | 159 886 | 129 | 50 | 79 | 84 | 81 | 55 | No | 88 152 | 84 | 20 465 | 13 | 25 634 | 16 |
| *Magnolia kwangsiensis* | | NC\_015892 | 159 667 | 129 | 51 | 78 | 84 | 81 | 52 | No | 88 030 | 84 | 18 971 | 13 | 26 333 | 16 |
| *Nymphaea alba* | | NC\_006050 | 159 930 | 130 | 50 | 80 | 85 | 48 | 55 | No | 90 014 | 85 | 19 562 | 13 | 25 177 | 16 |
| *Sabia Yunnanencis* | | NC\_029431 | 162 080 | 131 | 52 | 79 | 86 | 32 | 54 | No | 89 998 | 84 | 18 892 | 13 | 26 595 | 17 |
| *Annona cherimola* | | *KU563738* | 201 723 | 164 | 88 | 76 | 114 | 64 | 50 | No | 69 771 | 61 | 2 966 | 1 | 64 493 | 51 |
| *Annona cherimola* | |  | 163 883 | 137 | 56 | 81 | 88 | 32 | 56 | Yes | 89 724 | 80 | 18 654 | 13 | 27 750 | 22 |
| *Uvaria afzelii* | |  | 167 530 | 139 | 54 | 85 | 88 | 30 | 58 | Yes | 91 296 | 83 | 20 010 | 13 | 28 045 | 22 |
| *Annona muricata* | |  | 152 889 | 134 | 53 | 81 | 85 | 32 | 57 | Yes | 83 382 | 78 | 16 126 | 13 | 26 692 | 22 |
| *Anaxagorea phaeocarpa Mart. vel aff.* | |  | 162 223 | 136 | 53 | 83 | 86 | 29 | 57 | Yes | 88 881 | 81 | 19 503 | 13 | 26 903 | 21 |
| *Asimina pygmaea* | |  | 160 649 | 136 | 55 | 81 | 91 | 35 | 56 | Yes | 86 824 | 78 | 17 634 | 13 | 28 095 | 22 |
| *Cananga kirkii* | |  | 168 388 | 135 | 54 | 81 | 92 | 32 | 60 | Yes | 92 107 | 81 | 20 060 | 13 | 28 028 | 23 |
| *Cremastosperma leiophyllum* | |  | 163 483 | 138 | 55 | 83 | 93 | 33 | 60 | Yes | 89 325 | 81 | 19 237 | 13 | 27 460 | 22 |
| *Oxandra macrophylla* | |  | 169 521 | 139 | 55 | 84 | 93 | 33 | 60 | Yes | 93 592 | 83 | 20 085 | 13 | 31 063 | 22 |
| *Xylopia peruviana* | |  | 159 867 | 138 | 55 | 84 | 95 | 34 | 61 | Yes | 87 758 | 82 | 16 711 | 13 | 27 595 | 22 |
| \* Compared to the plastomes from *Magnolia kwangsiensis* and *Liriodendroon tulipifera*, the plastome of *Uvaria afzelii* has an inversion spanning over the genes atpE, atpB, rbcL, accD, psaI and ycf4. | | | | | | | | | | | | | | | | |

**Figure legends**

**Fig. 1** Percentageof mapped reads of *Uvaria afzelii* to different genomes. Genomes used to map the raw reads were the following: *Homo sapiens*, *Mus musculus*, *Rattus norvegicus*, *Drosophila melanogaster*, *Anopheles gambiae*, and *Saccharomyces cerevisiae*. The chloroplast genome used was from *Miliusa cuneata*. “No library” refers to all reads that did not map to any of the libraries used

**Fig. 2** Plastome map of *Uvaria afzelii*. Genes in the outer side of the circle have sense oriented genes, whereas the inner side shows anti-sense oriented genes. Genes have been coloured according to their function