

Chloroplast phylogenomic analyses resolve multiple origins of the *Kengyilia* species via independent polyploidization events

Shi-Yong Chen

Triticeae Research Institute

Hao Yan

Triticeae Research Institute

Li-Na Sha

Triticeae Research Institute

Ning Chen

Triticeae Research Institute

Yue Zhang

Triticeae Research Institute

Yi Wang

Triticeae Research Institute

Hou-Yang Kang

Triticeae Research Institute

Hai-Qin Zhang

Triticeae Research Institute

Yong-Hong Zhou

Triticeae Research Institute

Xing Fan (✉ fanxing9988@163.com)

Triticeae Research Institute

Research article

Keywords: Polyploid, Triticeae, multiple origins, maternal donor

Posted Date: February 5th, 2020

DOI: <https://doi.org/10.21203/rs.2.22669/v1>

License:   This work is licensed under a Creative Commons Attribution 4.0 International License.

[Read Full License](#)

Abstract

Background

Kengyilia is a group of allohexaploid species that arose from two hybridization events followed by genome doubling of three ancestral diploid species with different genomes **St**, **Y** and **P** in the wheat tribe. Estimating phylogenetic relationship in resolution of the maternal lineages has been difficult, owing to the extremely low rate of sequence divergence. Here, phylogenetic reconstructions based on the plastome sequences were used to explore the role of maternal progenitors in establishment of *Kengyilia* polyploid species.

Results

The plastome sequences of 11 *Kengyilia* species were analyzed together with 11 tetraploid species (**PP**, **StP**, and **StY**) and 33 diploid taxa representing 20 basic genomes in the Triticeae. Phylogenomic analysis and genetic divergence patterns suggested that (1) *Kengyilia* is closely related to *Roegneria*, *Pseudoroegneria*, *Agropyron*, *Lophopyrum*, *Thinopyrum*, and *Dasypyrum*; (2) both the **StY** genome *Roegneria* tetraploids and the **PP** genome *Agropyron* tetraploids severed as the maternal donor during the speciation of *Kengyilia* species; (3) the different *Kengyilia* species derived their **StY** genome from different *Roegneria* species.

Conclusion

Multiple origins of species via independent polyploidization events have occurred in the genus *Kengyilia*, resulting in a maternal haplotype polymorphism. This helps explain the rich diversity and wide adaptation of polyploid species in the genus *Kengyilia*.

Background

Polyploidy, defined as the possession of two or more sets of homologous chromosomes following whole-genome duplication, is a major mechanism in plant evolution and speciation [1, 2]. Recent studies even suggested that multiple origins (including independent origin) of polyploid species are the rule rather than the exception [3–6]. Polyploidy promotes variability through change in the chromosomal number per se, increased genetic diversity and genomic reorganization, leading to benefits in new phenotypes and evolutionary innovation in physiological and ecological flexibility [7, 8]. However, a clear and appropriate identification of phylogenetic relationships among taxa and genomics is needed [5].

Kengyilia Yen et J.L. Yang, a polyploid perennial genus in the wheat tribe (Poaceae: Triticeae), includes about 22 perennial species that distributed in a different range of natural habitats over the upper and middle mountain ranges of Central Asia and the Qinghai-Tibetan Plateau [9]. A comparison of the morphological features among the genus *Kengyilia*, *Roegneria*, *Elymus*, and *Agropyron* suggested that the species in *Kengyilia* is intermediate between species of *Roegneria* C. Koch and *Agropyron* Gaertn, but

with some distinct morphological divergence in spikelet characters between *Kengyilia* and *Roegneria* and between *Kengyilia* and *Agropyron* [10]. *Kengyilia* species exhibit variation with high (*K. grandiglumis*) to low (*K. thoroldiana*) plants, lax (*K. rigidula*) to dense (*K. hirsuta*) spikes, adnate (*K. longilumis*) to incohesive (*K. stenachyra*) spikelets to rachis, yellow (*K. gobicola*) to black (*K. melanthera*) anthers. All the species of *Kengyilia* are allohexaploids ($2n = 6x = 42$) with StYP genomes [9, 11]. The St and P genomes are originated from *Pseudoroegneria* (Nevski) Á Löve and *Agropyron* Gaertn., respectively [12]. It is unknown where the Y genome originates, although it is a fundamental *Kengyilia* genome [5, 11]. Cytogenetic evidence suggested that speciation of the *Kengyilia* polyploid was derived from hybridization between tetraploid *Roegneria* C. Koch ($2n = 4x = 28$, StY) and diploid *Agropyron* ($2n = 2x = 14$, P) species [5, 11]. Analysis of nuclear single-copy P_{gk1} gene sequences suggested that *Kengyilia* species from the Central Asia and the Qinghai-Tibetan plateau have independent origins with geographically differentiated P genome donors [13]. Data from chloroplast trnL-F [14], matK [15], rbcL [15], trnH-psbA [15], and mitochondrial CoxII [16] suggested that different species of *Kengyilia* have derived their maternal lineages from either the species of *Pseudoroegneria* or the species of *Agropyron* or unknown donor. Analysis of trnL-F [14] and matK and rbcL [15] sequences showed that four species of *Kengyilia* (*K. kokonorica*, *K. melanthera*, *K. mutica*, and *K. thoroldiana*) were related to species of *Agropyron*, and the remaining sampled species were close to species of *Pseudoroegneria*. In trnH-psbA tree, four species of *Kengyilia* (*K. grandiglumis*, *K. hirsuta*, *K. laxiflora*, and *K. rigidula*) were grouped with *Agropyron*, which is inconsistent with the maternal relationship presented by the trnL-F [14] and the matK and rbcL [15] sequence data. Moreover, analysis of CoxII suggested that some species of *Kengyilia* (e.g.: *K. batalinii*), *Agropyron*, and *Pseudoroegneria* formed a paraphyletic grade with zero-length branches [16]. While these studies added our understanding of phylogenetic relationship of *Kengyilia*, the molecular phylogenies based on published chloroplast DNA (trnL-F, matK, rbcL, trnH-psbA) and mitochondrial sequences in resolution of the maternal lineages of *Kengyilia* species are still in dispute due to either the unresolved gene tree with polytomies or incongruence among cytoplasmic gene data [14–16]. Moreover, the processes that have driven polyploid diversification and speciation, especially with regards to which tetraploid and diploid species as maternal progenitors were involved in hexaploid evolution in *Kengyilia*, remain unclear. Thus, to better understand the maternal contribution to the species of *Kengyilia*, it is essential to obtain a well comparative study of chloroplast genome-wide in *Kengyilia* and its relatives covering nearly all of the genomic combinations in Triticeae.

Here, integrating 38 newly and 18 previously sequenced plastomes representing the StYP genomes and its related tetraploid and diploid genomic types in Triticeae, this study applies phylogenetic reconstruction methods in combination with estimate of genetic distance among coding region to clarify maternal lineage relationships. Our objectives are to demonstrate a phylogenomic framework for illustrating the maternal donor of *Kengyilia* polyploids and to explore the role of maternal progenitors in establishment of *Kengyilia* polyploid.

Results

Characteristics of chloroplast genomes and genes

All sequenced genomes are very similar to published cp genomes of Triticeae [17, 18] and rather conservative in genome structure and gene content. Their genome size ranges from 134,985 in *R. ciliaris* to 135,489 bp in *A. cristatum* (ZY09064). All plastomes exhibited a typical quadripartite structure that included a pair of IRs separated by a large single copy region (LSC) and a small single copy region (SSC) and contained a total of 109 genes (including 76 protein coding genes, 29 tRNA genes and 4 rRNA genes). Assemblies in genus *Kengyilia* averaged 135,113 bp, with an estimated 0.064% insertion data (compared to *Pseudoroegneria libanotica* reference); genus *Roegneria* assemblies averaged just less than 135,079 bp (0.039% estimated insertion data, compared to *Pse. libanotica* reference).

Analysis of co-linearity is inferred for two diploid taxa representing St and P genomes, one tetraploid species with the StP genome, one tetraploid species with the StY genome, and three hexaploid species with the StYP genome (Fig. 1). Despite a high degree of co-linearity among these genomes due to the conservative in chloroplast genome structure and gene content, five big indels (at position 17819–18278 bp, 56172–56963 bp, 62664–63130 bp, 83590–84338 bp, 130804–131592 bp, respectively) were detected between the St- and P-containing lineages, which is an indicative of high genetic divergence between them.

The features of each of the 76 protein-coding gene in the diploid-polyploid plastome data are summarized in Table S1. The lengths of each gene ranged from 90 (*petN*) to 4,440 (*rpoC2*) bp. The proportion of variable sites (variable sites/total sites, V/T) varied from 0 (e.g. *petG*) to 3.36% (*rpl32*). The ratio of parsimony-informative characters per total aligned characters was greatest for *petL* (2.08%) and lowest for *petG*, *psbF*, and *rpl23* (0).

Phylogenetic analyses

Bayesian phylogenetic reconstruction of the plastome data under the GTR + G + I model resulted in a tree with high posterior probability support across most clades. ML analyses in IQ-Tree under the TVM + F + R3 model recovered the same topology as the Bayesian analyses. The tree illustrated in Fig. 2 was the BI tree with statistic supports (UFboot, SH-aLRT, and PP) above branches. The phylogenetic tree showed that the plastome sequences of *Kengyilia* were split into two major clades (Clade I and II) with consistent statistical support (100% UFboot and SH-aLRT; 1.0 PP). The Clade I included *Thinopyrum* (E^b), *Lophopyrum* (E^e) \square *Dasypyrum* (V), *Pseudoroegneria* (St), and all the sampled St-containing (*Douglasdeweya*, StP; *Roegneria*, StY; *Kengyilia*, StYP) polyploid species (except for *Kengyilia melanthera*) with consistent statistical support (100% UFboot and SH-aLRT; 1.0 PP). In this clade, *Thinopyrum* \square *Lophopyrum* \square *Dasypyrum*, *Douglasdeweya*, four species of *Pseudoroegneria* (*Pse. stipifolia*, *Pse. cognata*, *Pse. libanotica* and *Pse. tauri*), two species of *Roegneria* (*R. grandis* and *R. ciliaris*), and four species of *Kengyilia* (*K. alata*, *K. hirsuta*, *K. laxiflora*, and *K. batalinii*) were in one subclade (99.8% UFboot, 100% SH-aLRT, and 1.0 PP). *Kengyilia alata* from central Asia formed a paraphyletic grade with *Thinopyrum* \square *Lophopyrum* \square *Dasypyrum*, *Douglasdeweya*, and two species of *Pseudoroegneria* (*Pse. stipifolia* and *Pse. cognata*). Two *Kengyilia* species from central Asia (*K. hirsuta* and *K. batalinii*) and one *Kengyilia* species from the Qinghai-Tibetan Plateau (*K. laxiflora*) were clustered with two species

of *Roegneria* (*R. grandis* and *R. ciliaris*) (95% UFboot, 95% SH-aLRT, and 1.0 PP). *Kengyilia kokonorica* from the Qinghai-Tibetan Plateau and two species of *Pseudoroegneria* (*Pse. libanotica* and *Pse. tauri*) formed a paraphyletic grade in the subclade. Five species of *Kengyilia* from the Qinghai-Tibetan Plateau (*K. thoroldiana*, *K. grandiglumis*, *K. mutica*, *K. stenachyra*, and *K. rigidula*) were grouped with one species of *Roegneria* (*R. longearistata*) (100% UFboot, 100% SH-aLRT, and 1.0 PP), and this group sister to two accessions of *Pseudoroegneria spicata* (99.6% UFboot, 100% SH-aLRT, and 1.0 PP). The clade II contained all sampled *Agropyron* species (*A. cristatum* and *A. mongolicum*) and *K. melanthera* from the Qinghai-Tibetan Plateau (100% UFboot, 100% SH-aLRT, and 1.0 PP).

Statistic of K2-p distance matrix

A distance matrix including 1,664 genetic values was generated to investigate the relationship between the plastomes of *Kengyilia* and those of its closely relatives (Table S2). The Hopkins statistic was found to be 0.2057, indicating that the data is highly clusterable (Fig. 3A). Analysis of both bivariate cluster and clustering dendrogram based on the method of the hierarchical agglomerative clustering shows four major clusters (Fig. 3B and 3C), which correspond to the four genomic types (P/StYP, E^e/E^b, StY, and St/StP/StY/StYP). This is also well congruent with the groupings in phylogenomic tree inferred from the plastome data including all sampled Triticeae plants. The first cluster included all sampled *Agropyron* species (*A. cristatum*, ACR; *A. mongolicum*, AMO) and *K. melanthera* (KME). The second cluster contained one species of *Roegneria* (*R. longearistata*, RLO) and five species of *Kengyilia* (*K. stenachyra*, KST; *K. rigidula*, KRI; *K. grandiglumis*, KGR; *K. mutica*, KMU; *K. thoroldiana*, KTH). The third cluster consisted of *Thinopyrum* (TBE) and *Lophopyrum* (LEL). The fourth cluster comprised all the sampled *Pseudoroegneria* (*Pse. spicata*, PSP; *Pse. stipifolia*, PST; *Pse. cognata*, PCO; *Pse. libanotica*, PLI; *Pse. tauri*, PTA), *Douglasdeweya* (DDE), two species of *Roegneria* (*R. grandis*, RGR; *R. ciliaris*, RCI), and four species of *Kengyilia* (*K. hirsuta*, KHI; *K. laxiflora*, KLA; *K. alata*, KAL; *K. batalinii*, KBA).

Discussion

The cpDNA-based (*trnL-F*, *matK*, *rbcl*, and *trnH-psbA*) phylogeny of the genus *Kengyilia*, especially with regard to the origin of maternal donor during hexaploid polyploidization events, were largely unresolved due to the occurrence of many polytomies and incongruence among published gene tree [14, 15]. Ma et al. [19] pointed out that despite missing samples, phylogenetic analysis of plastome sequences can offer the greatest phylogenetic resolution. In this study, a resolved tree with highly statistic support was inferred from the plastome sequences of *Kengyilia* and those of its relatives in Triticeae, allowing the relationship regarding to the maternal lineages of *Kengyilia* to be clarified.

In phylogenomic tree, ten species of *Kengyilia* (*K. alata*, *K. hirsuta*, *K. laxiflora*, *K. batalinii*, *K. kokonorica*, *K. thoroldiana*, *K. grandiglumis*, *K. mutica*, *K. stenachyra*, and *K. rigidula*), *Roegneria*, and *Pseudoroegneria* were in one group with consistent support, indicating that *Pseudoroegneria* is likely to be the maternal donor of these ten StYP genome *Kengyilia* species and the sampled StY genome *Roegneria* species. Since *Kengyilia* species arose from two hybridization events followed by three

genome doublings (the St, Y and P genomes), with one firstly generating the StY genome *Roegneria* and the other forming the StYP genome *Kengyilia* [11, 13], *Roegneria* severed as the maternal donor during the speciation of the ten *Kengyilia* species.

Analysis of *trnL-F* suggested that four species of *Kengyilia* (*K. kokonorica*, *K. melanthera*, *Kengyilia mutica*, and *Kengyilia thoroldiana*) were closely related to species of *Agropyron* [14]. A similar deep-level relationships regarding to maternal lineages is also presented by Luo et al. [15], although molecular characters (including *matK*, *rbcL*, and *trnH-psbA*) and more taxa were sampled from *Kengyilia*. In this study, only *K. melanthera* was grouped with the species of *Agropyron*, and the remaining three species (*K. kokonorica*, *K. mutica*, and *K. thoroldiana*) were placed into the clade including St-containing species. Moreover, the plastome sequence of *K. melanthera* and *Agropyron* are obviously distinct from those of the St-containing species. Thus, the molecular phylogenies based on published cpDNA fragments and the present plastome sequence data in resolution of the placement of *K. kokonorica*, *K. mutica*, and *K. thoroldiana* led to apparently contradictory results. Discordances among phylogenetic trees result from methodological artifacts (e.g., sampling error and/or a failure of molecular characters) and the complex dynamics of the evolutionary processes in organisms (e.g., hybridization and/or ancestral polymorphisms) [6, 20]. Sampling error is likely to be the candidate for the current incongruences because our samples for the comparative phylogenies with *Kengyilia* species included nearly all of the monogenomic genera accepted in genome-based classifications of the Triticeae, and most monogenomic genera were not covered in previous study [14, 15]. It is well known that molecular characters can affect the accuracy of phylogenetic estimates [19]. Incongruences would also be the result of lack of molecular characters. Less molecular characters in cpDNA regions, as indicated by our estimate for the variable features of each chloroplast protein-coding genes (Table S2), together with its slowly evolving rates in chloroplast genome, would not only provide a few variable information for the accuracy of phylogenetic reconstruction but also result in the occurrence of polytomies in phylogenetic tree. On the contrary, the plastome data offer enough molecular characters for the accuracy of phylogenetic estimates with well-supported topology. Both hybridization and ancestral polymorphisms acting alone or in concert can generate discordance and therefore are the principal processes to explain the phylogenetic incongruence in Triticeae species [6, 21]. Analysis of genetic distance matrix based on the 52 protein-coding genes suggested that *Lophopyrum* and *Thinopyrum* are closely related to the St-containing species. In phylogenomic tree inferred from complete chloroplast genome, *Lophopyrum*, *Thinopyrum*, *Dasypyrum*, and two species of *Pseudoroegneria* (*Pse. stipifolia* and *Pse. congata*) form a monophyletic group. These results indicated *Lophopyrum*, *Thinopyrum*, *Dasypyrum*, *Pseudoroegneria* (most likely *Pse. stipifolia* and *Pse. congata*) shared ancestral polymorphisms due to incomplete diversification of common maternal ancestry. Such ancestral polymorphisms could be genetically transmitted to some polyploid species (e.g.: StP, StY, StYP) via the hybridization between *Pseudoroegneria* as female parent and the donors with Y and/or P genomes. The hypothesis of hybridization is also a likely candidate to explain the conflict because different polyploid species with the same genotypes could derive from different parental donors via independent hybridization events, generating a diverse array of polyploid genotypes in Triticeae [5, 22]. The present plastome data also provides support for the independent origin

some polyploid species, which can be shown by different *Kengyilia* species that was grouped with different *Roegneria* species in a phylogenetic tree. For example, in the clade I of phylogenomic tree, three *Kengyilia* species (*K. hirsuta*, *K. laxiflora*, and *K. batalinii*) were clustered with *R. grandis* with strongly statistic support (100% UFboot, 100% SH-aLRT, and 1.0 PP), and five *Kengyilia* species (*K. thoroldiana*, *K. grandiglumis*, *K. mutica*, *K. stenachyra*, and *K. rigidula*) were grouped with *R. longearistata* (100% UFboot, 100% SH-aLRT, and 1.0 PP). Analysis of genetic distances based on 52 protein-coding sequences also presented similar results. Sympatric distribution among *R. grandis*, *R. longearistata* and *Agropyron* species have provide an opportunity in physical proximity for hybridization events. It is thus suggested that the different *Kengyilia* species derived their StY genome from different *Roegneria* species. Our data also indicated that *Agropyron* species severed as the maternal donor during the speciation of *K. melanthera*, providing additional support for the independent origin of different *Kengyilia* species. However, it seems unlikely that the maternal *Agropyron* lineage in *K. melanthera* resulted from hybridization between high ploidy *Roegneria* species with StY genomes (served as paternal donor) and diploid P genome *Agropyron* species. One possible explanation is that the P genome of *K. melanthera* originated from the tetraploid *Agropyron* lineage as the female parent. Given the present data, multiple origins of polyploid species result in a maternal haplotype polymorphism and could explain the rich diversity and wide adaptation of polyploid species in the genus *Kengyilia* [11].

Conclusions

The present analysis of phylogenetic relationships in *Kengyilia* based the plastome sequences revealed that both *Roegneria* and *Agropyron* tetraploid species severed as the maternal donor during the speciation of *Kengyilia* species, and different *Roegneria* species contributed their StY genome to different *Kengyilia* species. This is an indicative of independent origin of different *Kengyilia* species, which shed new light on our understanding of the maternal lineages, polyploidization events and speciation process of *Kengyilia*.

Methods

Plant materials

A total of 22 polyploids, comprising 11 *Kengyilia* (StYP genomes) species, seven *Agropyron* (PP) tetraploids, one *Douglasdewya* (StP genomes) species, and three *Roegneria* (StY genomes) species, were analyzed together with 33 diploid taxa representing 20 basic genomes in the Triticeae (Table S3). *Brachypodium distachyon* was used as outgroup. The chloroplast genome sequences of *Triticum-Aegilops* complex, *Secale*, *Pseudoroegneria cognata*, *Pseudoroegneria stipifolia*, *Pseudoroegneria strigosa*, and *Pseudoroegneria tauri* were from the published data [17, 18]. The remaining sequences in Table S3 in Triticeae were newly sequenced. Sample information including accession numbers, origins, genome type, ploidy, and GenBank accession data were also listed in Table S3. The seed materials with PI and W6 numbers were kindly provided by American National Plant Germplasm System (Pullman, Washington, USA). The seed materials with ZY and Y numbers were collected from the Triticeae species

that are not endangered, and they grow in public area where no permission and formal identification for collection is needed in China. The plants used for sequencing and voucher specimens are deposited at Herbarium of Triticeae Research Institute, Sichuan Agricultural University, China (SAUTI).

Plastome sequencing and data assembly

DNA extractions were performed on young green leaf material dried in silica using the DNeasy Plant Mini kit (Qiagen, Valencia, USA) according to the manufacturer's instructions after a liquid nitrogen homogenization step. The plastomes were amplified in overlapping fragments using the long-range PCR method of Yang et al. [23]. The PCR products was fragmented into short insert (400–600 bp) to construct the sequencing paired-end library according to the NEBNext® protocol. DNA from each individual was indexed using tags and pooled together in one lane of an Illumina HiSeq 4000 PE150 for sequencing. The sequenced plastomes were assembled and annotated following Yang et al. (2014). The plastome of *Pseudoroegneria libanotica* (GenBank accession number KX822019) and *Agropyron cristatum* (GenBank No. KY126307) were used as the reference genomes for assembly of all new sequenced accessions.

Sequence alignment and analysis

The complete chloroplast sequences were aligned with MAFFT v. 7 [24] using the default settings. All alignments were visually inspected in MEGA 6.0 [25] and manually adjusted where needed. We also conducted a co-linear analysis using the software LASTZ, and the results were visualized using AliTV [26]. Multiple alignment of protein-coding sequence was conducted using ClustalW in MEGA 6 [25], with manual adjustment. Amino acid translations were used to guide the nucleotide alignments. The sequence statistics, including nucleotide substitutions, Kimura 2-parameter (K2-p) distances, transition/transversion ratio, and variability of the sequences, were calculated by MEGA 6 [25].

To estimate the genetic differentiation of protein-coding sequences between *Kengyilia* and its closely relatives, K2-p model was used to calculate the genetic distances of each protein-coding sequences of fifty-two genes (76 unique genes, excluding 24 which have no variable nucleotide sites and/or are < 200 bp) between *Pasthyrostachys* as the outgroup and the samples including *Kengyilia* and its closely relatives. A total of 1,664 (32 samples and 52 protein-coding genes) genetic distances were used to estimate genome relationship using hierarchical agglomerative clustering in R (Version 3.4.2; Vienna, Austria). Hopkins statistic was used for the evaluation of clustering tendency. The optimal number of clusters was determined by "fviz_dend" algorithm in the R package 'factoextra' Version 1.0.5. Bivariate cluster (k-means clustering) analysis based on genetic distances and agglomerative hierarchical clustering was performed by the 'clustplot' function in the R package "cluster" package (Version 2.0.6).

Phylogenetic analysis

Because complete chloroplast genome sequences offer the greatest phylogenetic resolution [19], phylogenomic trees were generated from all sampled complete chloroplast genomes. Phylogenomic analyses were conducted using maximum likelihood (ML) and Bayesian inference (BI). ML analysis was

performed using the IQ-Tree software (<http://www.cibiv.at/software/iqtree/>). The TVM + F + R3 model was selected for ML analysis according to ModelTest implemented in IQ-Tree. To assess branch support, the IQ-Tree analyses used the ultrafast bootstrap approximation (UFboot) with 10000 replicates [27] and the SH-like approximate likelihood ratio test (SH-aLRT) with 1000 bootstrap replicates [28].

The evolutionary model used for BI analysis was determined using ModelTest v3.7 [29] with Akaike information criterion (AIC). BI analysis was performed using MrBayes v3.2 [30] under the GTR + G + I model that was identified as the best fit by ModelTest. Four MCMC (Markov Chain Monte Carlo) chains (one cold and three heated), applying MrBayes default heating values ($t = 0.2$), were run 1,000,000 generations for plastome data, with each sampled in each data set every 100 generations. Fifty percent of majority consensus trees were generated with a relative burn-in of 25%. The statistical confidence in nodes was estimated by posterior probabilities (PP). A PP-value less than 90% was not included in figures.

Abbreviations

cp, chloroplast; LSC, large-single-copy; SSC, small-single-copy; UFBoot, ultrafast bootstrap; SH-aLRT, Shimodaira–Hasegawa approximate likelihood ratio test; PP, posterior probabilities.

Declarations

Acknowledgements

We are grateful to American National Plant Germplasm System (Pullman, Washington, USA) providing the part seed materials for this study.

Funder

The Funder is the National Natural Science Foundation of China (Nos. 31870360, 31470305). The Funder had no role in study design or manuscript writing.

Availability of data and materials

The sequencing data from our study was deposited in the National Center for Biotechnology Information (NCBI).

Authors Contributions

XF and YHZ designed the research. SYC, HY, LNS, NC, and YZ performed the research. YW, HYK, and HQZ collected plant materials. XF, SYC, and HY analyzed the data and wrote the manuscript. XF and YHZ edited the manuscript. All authors read and approved the final manuscript.

Ethics approval and consent to participate

Three used samples, including *Amblyopyrum muticum*, *Aegilops sharonensis*, and *Aegilops bicornis*, are listed as endangered species on the red list. We declared that we did not collect the three endangered species from any field, and we just used the chloroplast genome sequences from the NCBI public web where no permission for download is needed.

Consent for publication

Not applicable.

Competing interests

The authors declare that they have no competing interests.

References

1. Otto SP. The evolutionary consequences of polyploidy. *Cell*. 2007; 131: 452–462.
2. Spoelhof JP, Soltis PS, Soltis DE. Pure polyploidy: closing the gaps in autopolyploid research. *J Syst Evol*. 2017; 55: 340-352.
3. Soltis PS, Soltis DE. The role of genetic and genomic attributes in the success of polyploids. *Proc Natl Acad Sci USA*. 2000; 97: 7051–7057.
4. Symonds VV, Soltis PS, Soltis DE. Dynamics of polyploid formation in *Tragopogon* (Asteraceae): recurrent formation, gene flow, and population structure. *Evolution*. 2010; 64: 1984–2003.
5. Fan X, Sha LN, Dong ZZ, Zhang HQ, Kang HY, Wang Y, Wang XL, Zhang L, Ding CB, Yang RW, Zheng YL, Zhou YH. Phylogenetic relationships and Y genome origin in *Elymus* L. sensu lato (Triticeae; Poaceae) based on single-copy nuclear *Acc1* and *Pgk1* gene sequences. *Mol Phylogenet Evol*. 2013; 69: 919-928.
6. Sha LN, Fan X, Li J, Liao JQ, Zeng J, Wang Y, Kang HY, Zhang HQ, Zheng YL, Zhou YH. Contrasting evolutionary patterns of multiple loci uncover new aspects in the genome origin and evolutionary history of *Leymus* (Triticeae; Poaceae). *Mol Phylogenet Evol*. 2017a; 114: 175-188.
7. Ramsey J, Ramsey TS. Ecological studies of polyploidy in the 100 years following its discovery. *Philos T R Soc B*. 2014; 369: 20130352.
8. Sha LN, Fan X, Wang XL, Dong ZZ, Zeng J, Zhang HQ, Kang HY, Wang Y, Liao JQ, Zhou YH. Genome origin, historical hybridization and genetic differentiation in *Anthosachne australasica* (Triticeae; Poaceae), inferred from chloroplast *rbcl*, *trnH-psbA* and nuclear *Acc1* gene sequences. *Ann Bot*. 2017b; 119: 95-107.
9. Yang JL, Yen C, Baum BR. *Kengyilia*: synopsis and key to species. *Hereditas*. 1992; 116: 25-28.
10. Baum BR, Yang JL, Yen C. Taxonomic separation of *Kengyilia* (Poaceae: Triticeae) in relation to nearest related *Roegneria*, *Elymus*, and *Agropyron*, based on some morpho- logical characters. *Plant Syst Evol*. 1995; 194: 123–132.

11. Yen C, Yang JL, Baum BR. Biosystematics of Triticeae. 2006; Volume 3. Chinese Agriculture Press, Beijing.
12. Löve A. Conspectus of Triticeae. Feddes Rep. 1984; 95: 425–524.
13. Fan X, Sha LN, Zeng J, Kang HY, Zhang HQ, Wang XL, Zhang L, Yang RW, Ding CB, Zheng YL, Zhou YH. Evolutionary dynamics of the *Pgk1* gene in the polyploid genus *Kengyilia* (Triticeae: Poaceae) and its diploid relatives. PLoS One, 2012; 7: e31122.
14. Zhang C, Fan X, Yu HQ, Wang XL, Zhou YH. Different maternal genome donor to *Kengyilia* (Poaceae: Triticeae) species inferred from chloroplast *trnL-F* sequences. Biol Plant. 2009; 53: 759–763.
15. Luo X, Tinker NA, Fan X, Zhang HQ, Sha LN, Kang HY, Ding CB, Liu J, Zhang L, Yang RW, Zhou YH. 2012. Phylogeny and maternal donor of *Kengyilia* species (Poaceae: Triticeae) based on three cpDNA (*matK*, *rbcL* and *trnH-psbA*) sequences. Biochem Syst Ecol. 2012; 44: 61-69.
16. Zeng J, Fan X, Zhang L, Wang XL, Zhang HQ, Zhou YH. Molecular phylogeny and maternal progenitor implication in the genus *Kengyilia* (Triticeae: Poaceae): evidence from *coxII* intron sequences. Biochem Syst Ecol. 2010; 38: 202–209.
17. Gornicki P, Zhu H, Wang J, Challa GS, Zhang Z, Gill BS, Li W. The chloroplast view of the evolution of polyploid wheat. New Phytol. 2014; 204: 704-714.
18. Bernhardt N, Brassac J, Kilian B, Blattner FR. Dated tribe-wide whole chloroplast genome phylogeny indicates recurrent hybridizations within Triticeae. BMC Evol Biol. 2017; 17: 141.
19. Ma PF, Zhang YX, Zeng CX, Guo ZH, Li DZ. Chloroplast phylogenomic analyses resolve deep-level relationships of an intractable bamboo tribe Arundinarieae (Poaceae). Syst Biol. 2014; 63: 933-950.
20. Betancur-R R, Naylor GJP, Orti G. Conserved genes, sampling error, and phylogenomic inference. Syst Biol. 2014; 63: 257–262.
21. Mason-Gamer RJ. Phylogeny of a genomically diverse group of *Elymus* (Poaceae) allopolyploids reveals multiple levels of reticulation. PLoS ONE. 2013; 8: e78449.
22. Sun GL, Ni Y, Daley T. Molecular phylogeny of *RPB2* gene reveals multiple origin, geographic differentiation of H genome, and the relationship of the Y genome to other genomes in *Elymus* species. Mol Phylogenet Evol. 2008; 46: 897–907.
23. Yang JB, Li DZ, Li HT. Highly effective sequencing whole chloroplast genomes of angiosperms by nine novel universal primer pairs. Mol Ecol Resour. 2014; 14: 1024-1031.
24. Katoh K, Standley DM. MAFFT multiple sequence alignment software version 7: improvements in performance and usability. Mol Biol Evol. 2013; 30: 772-780.
25. Tamura K, Stecher G, Peterson D, Filipowski A, Kumar S. MEGA6: molecular evolutionary genetics analysis version 6.0. Mol Biol Evol. 2013; 30: 2725-2729.
26. Ankenbrand MJ, Hohlfeld S, Hackl T, Förster F. AliTV—interactive visualization of whole genome comparisons. PeerJ Comput Sci. 2017; 3: e116.
27. Minh BQ, Nguyen MAT, von Haeseler A. Ultrafast approximation for phylogenetic bootstrap. Mol Biol Evol. 2013; 30: 1188-1195.

28. Guindon S, Dufayard JF, Lefort V, Anisimova M, Hordijk W, Gascuel O. New algorithms and methods to estimate maximum-likelihood phylogenies: assessing the performance of PhyML 3.0. *Syst Biol.* 2010; 59: 307–321.
29. Ronquist F, Teslenko M, van Der Mark P, Ayres DL, Darling A, Höhna S, Larget B, Liu L, Suchard MA, Huelsenbeck JP. MrBayes 3.2: efficient Bayesian phylogenetic inference and model choice across a large model space. *Syst Boil.* 2012; 61: 539–542.
30. Posada D, Crandall KA. MODELTEST: testing the model of DNA substitution. *Bioinformatics.* 1998; 14:817–818.

Supplementary Information

Table S1. Features of each of the 76 protein-coding genes in analysis of polyploid and its diploid relatives.

Table S2. A distance matrix including 1,664 genetic values calculated from 59 protein-coding genes.

Table S3. List of taxa used in this study.

Figures

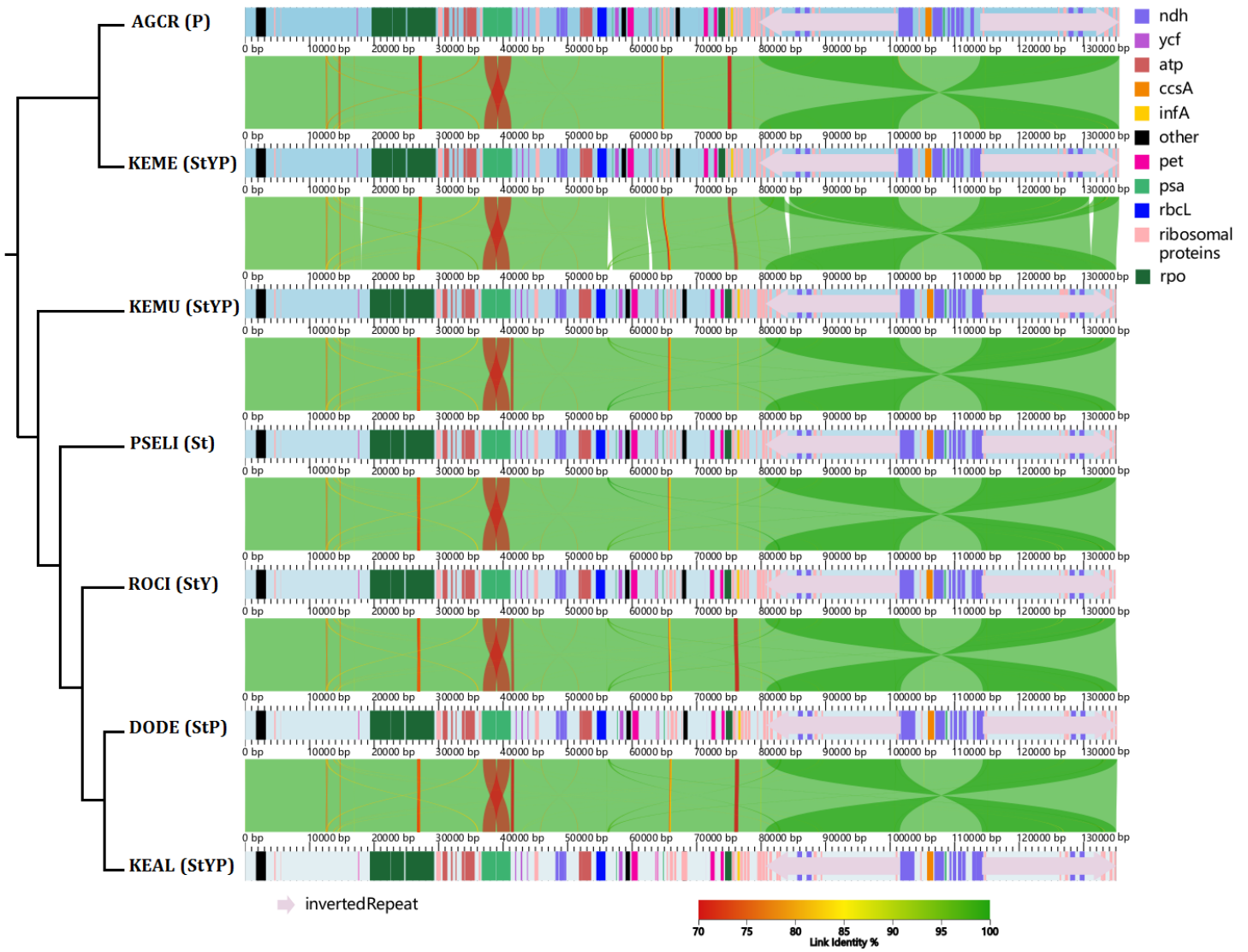


Figure 1

Co-linear analysis of the chloroplast genomes of *Kengyilia* species and its closely relatives.

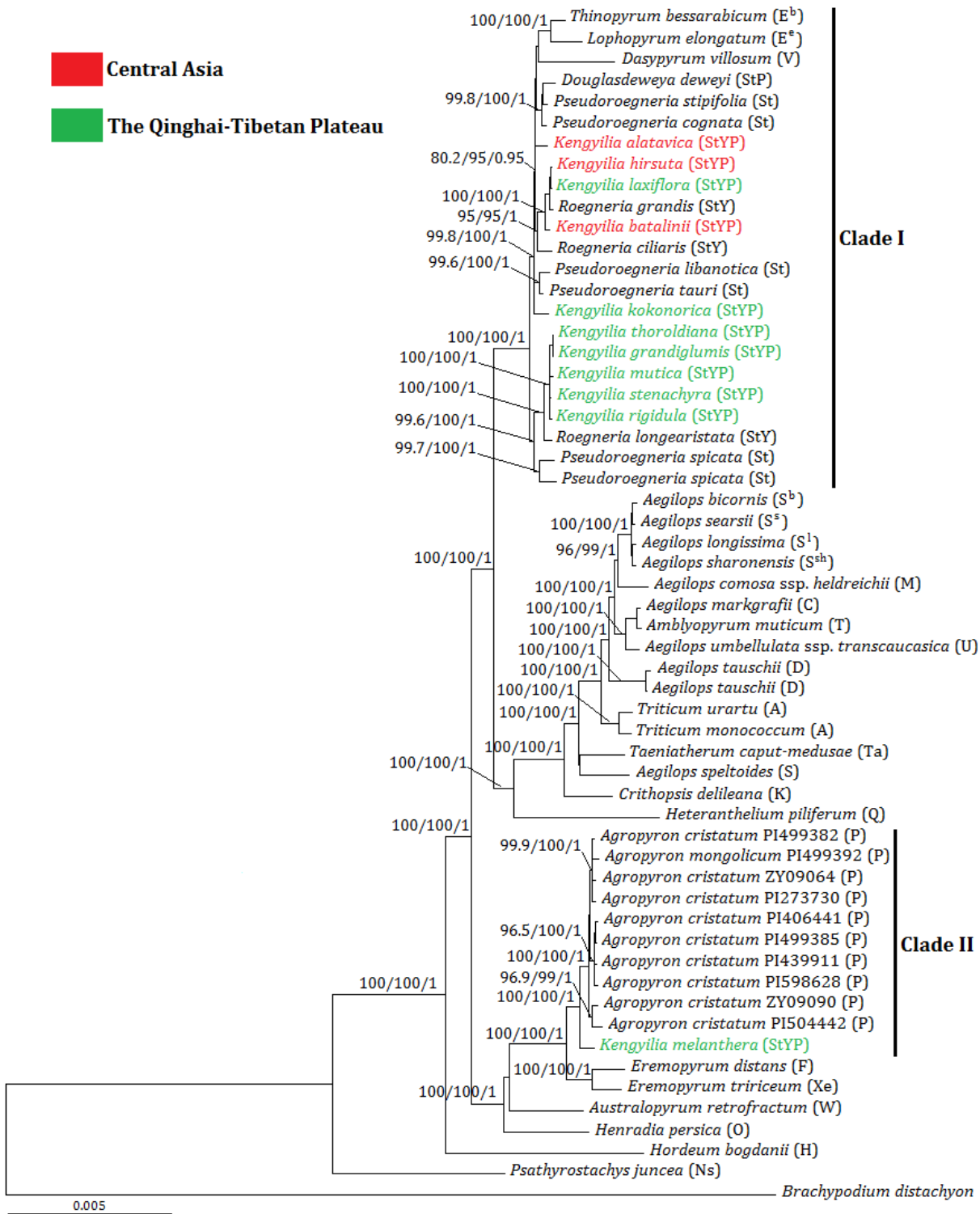


Figure 2

Bayesian tree inferred from the whole complete cp genome sequences for *Kengyilia* species and its diploid relatives in the Triticeae. Numbers above branches are the values of statistic support values, and values are indicated only if deemed robust as follows: UFboot \geq 95%/SH-aLRT \geq 80%/PP \geq 0.9. The capital letters in bracket indicate the genome type of the species. Different color labeled the geographic information *Kengyilia* species.

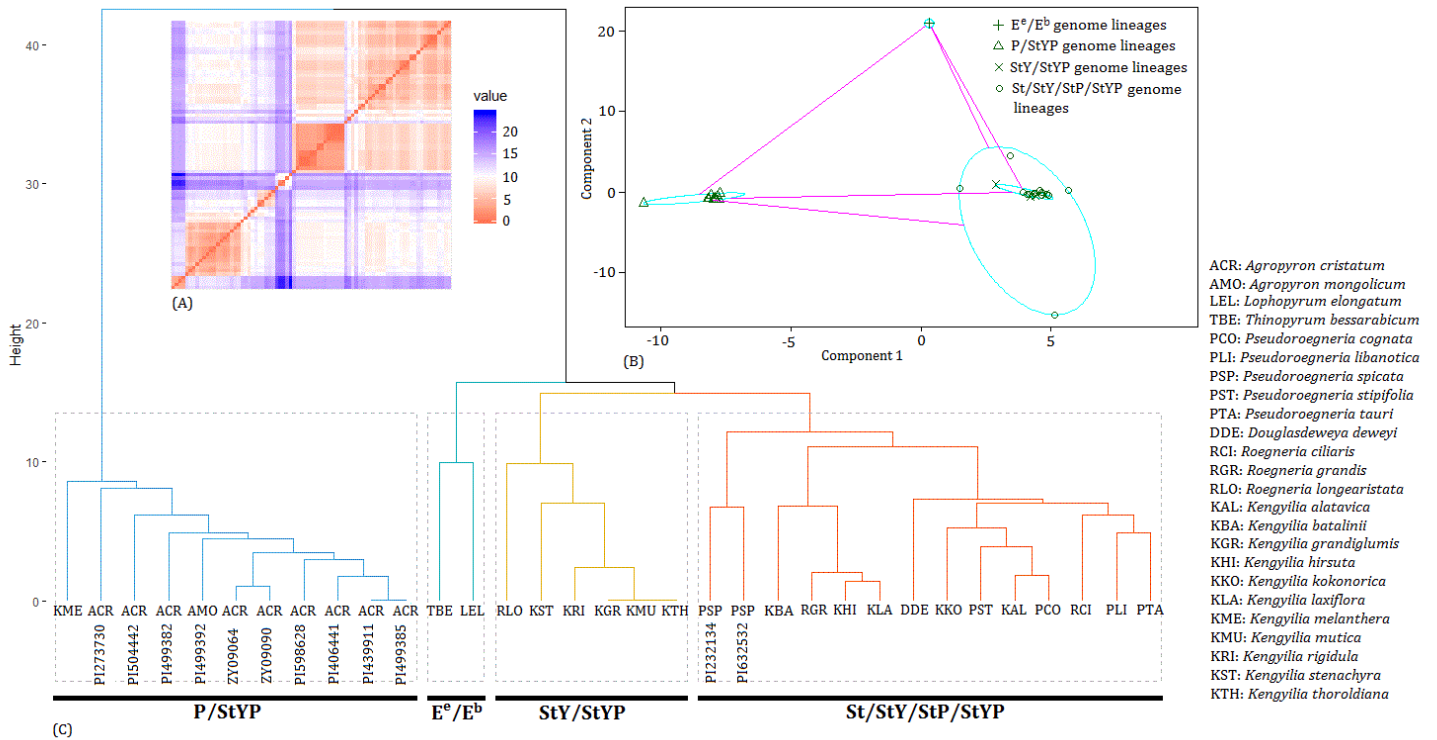


Figure 3

Cluster analysis based on a distance matrix including 1,664 genetic values calculated from 52 chloroplast protein-coding genes of three sampled *Kengyilia* specie and its closely relatives. (A) The Hopkins statistics. (B) The bivariate cluster plot based on the method of the hierarchical agglomerative clustering. (C) The clustering dendrogram based on the method of the hierarchical agglomerative clustering.

Supplementary Files

This is a list of supplementary files associated with this preprint. Click to download.

- [TableS2.xls](#)
- [TableS3.doc](#)
- [TableS1.xls](#)