Comparative genomics of 11 complete chloroplast genomes of Senecioneae (Asteraceae) species: DNA barcodes and phylogenetics

Andrew Wanyoike Gichira$^{1,2,3†}$, Sheila Avoga$^{1,2,3†}$, Zhizhong Li$^{1,2}$, Guangwan Hu$^{3,4}$, Qingfeng Wang$^{3,4*}$ and Jinming Chen$^{1*}$

Abstract

**Background:** Majority of the species within Senecioneae are classified in Senecio, making it the tribe's largest genus. Certain intergeneric relationships within the tribe are vaguely defined, with the genus Senecio being partly linked to this ambiguity. Infrageneric relationships within Senecio remain largely unknown and consequently, the genus has undergone continuous expansion and contraction over the recent past due to addition and removal of taxa. Dendrosenecio, an endemic genus in Africa, is one of its segregate genera. To heighten the understanding of species divergence and phylogeny within the tribe, the complete chloroplast genomes of the first five Senecio and six Dendrosenecio species were sequenced and analyzed in this study.

**Results:** The entire length of the complete chloroplast genomes was ~150 kb and ~151 kb in Dendrosenecio and Senecio respectively. Characterization of the 11 chloroplast genomes revealed a significant degree of similarity particularly in their organization, gene content, repetitive sequence composition and patterns of codon usage. The chloroplast genomes encoded an equal number of unique genes out of which 80 were protein-coding genes, 30 transfer ribonucleic acid, and four ribosomal ribonucleic acid genes. Based on comparative sequence analyses, the level of divergence was lower in Dendrosenecio. A total of 331 and 340 microsatellites were detected in Senecio and Dendrosenecio, respectively. Out of which, 25 and five chloroplast microsatellites (cpSSR) were identified as potentially valuable molecular markers. Also, through whole chloroplast genome comparisons and DNA polymorphism tests, ten divergent hotspots were identified. Potential primers were designed creating genomic tools to further molecular studies within the tribe. Intergeneric relationships within the tribe were firmly resolved using genome-scale dataset in partitioned and unpartitioned schemes. Two main clades, corresponding to two subtribes within the Senecioneae, were formed with the genus Ligularia forming a single clade while the other had Dendrosenecio, Pericallis, Senecio and Jacobaea. A sister relationship was revealed between Dendrosenecio and Pericallis whereas Senecio, and Jacobaea were closely placed in a different clade.
**Conclusion:** Besides emphasizing on the potential of chloroplast genome data in resolving intergeneric relationships within Senecioneae, this study provides genomic resources to facilitate species identification and phylogenetic reconstructions within the respective genera.

**Keywords:** Senecio, Dendrosenecio, Endemic, Chloroplast genome, Codon usage, Microsatellites, DNA barcodes, Phylogenomics

**Background**

Senecioneae, the largest tribe in the family Asteraceae, has over 160 genera with more than 3000 species, and new genera continue to be added (Chen et al. 2011; Nordenstam et al. 2009). The tribe is prominent for its size, and rich morphological and ecological diversity. It is mostly dominated by annual and perennial herbs, while the rest constitute shrubs, vines, trees, and epiphytes. It has a near cosmopolitan distribution, with southern Africa being one of its key diversity hotspot zones (Pelser et al. 2007). Majority of the species in the tribe are placed in Senecio L., making it one of the largest genera of angiosperms, with over 1250 species (Nordenstam et al. 2009). Senecio is characteristically diverse in morphology, life-history, growth forms, and thus, it has been markedly linked to the incongruous phylogenetic relationships within the tribe (Pelser et al. 2007). Its members are generally distinguished by style branches truncate with short sweeping hairs, separated stigmatic lines and sometimes with a median hair pencil, and with ecaudate anther bases and balusterform filament collar (Nordenstam 2007; Pelser et al. 2007).

Over the years, the genus has been under constant re-evaluation and reclassification, and until now, comprehensive infrageneric relationships are yet to be established. Consequently, numerous species have in the past been segregated as new genera mostly based on morphological, anatomical, and chromosomal variations (Jeffrey and Chen 1984; Jeffrey et al. 1977). One of such segregate genera is Dendrosenecio (Hauman ex Hedb.) B. Nord., upgraded by Nordenstam (1978) to constitute the Afromontane pachycaul taxa. Dendrosenecio was initially classified in Senecio based on the striking similarities in floral characters. It is therefore not surprising that the elevation of Dendrosenecio was at first controversial (Jeffrey et al. 1977) as the genus exhibited substantial morphological resemblances to other African perennials of Senecio. Despite these remarkable morphological similarities, amplified fragment length polymorphism analysis revealed considerable divergence between Senecio and Dendrosenecio (Knox and Palmer 1995b). Afterwards, internal transcribed spacer (ITS) data identified Oresbia Cron & B. Nord. as the closely related genus to Dendrosenecio (Pelser et al. 2007).

Majority of the segregated groups are now accepted on the basis of molecular data obtained from markers such as ITS (Pelser et al. 2007). However, it is evident that more valuable diagnostic molecular sequences are needed to further understand the generic and intergeneric relationships in Senecioneae. The large number of species, considerable variation in species life-history and over-dependence on morphological characters, the majority of which overlap, have been pointed out as the causes of the systematic conflict observed within Senecio. Similar to Senecio, infrageneric relationships within Dendrosenecio are still debatable, especially in relation to specific and subspecific classifications. Species of Dendrosenecio exhibit ‘mosaic of morphological variation’ arising from divergence and convergence as they dispersed to various geographical regions with similar habitat conditions (Knox 2005; Mabberley 1973). Besides, frequent hybridization events between species within each genus have been evidenced resulting in allopolyploid species (Hedberg 1957; Hegarty et al. 2012; Milton 2009). It is therefore imperative that more molecular markers and divergent regions are identified to facilitate species identification, speciation and adaptive evolution studies on species of Senecio and Dendrosenecio.

Partial plastid markers, species-specific or universal, have in the past decades been used to resolve phylogenetic relationships and species delimitations. This inclination is progressively being substituted by the use of plastid genome-scale data, resulting in improved phylogenetic resolutions and detailed evolutionary information about species at all taxonomic levels. Typically chloroplast DNA is uniparentally, maternally in angiosperms and paternally in gymnosperms, inherited and exhibits homologous recombination (Marechal and Brisson 2010). This attribute can greatly benefit studies on taxa that are affected by hybridization, introgression and convergent evolution. Additionally, chloroplast genomes are justifiably conserved in terms of gene composition and arrangement permitting comparative genomics even at the generic level. However, they harbour key variations e.g., in the inverted repeat (IR) size and positioning of the IR junctions even among close relatives (Downie and Jansen 2015), and in specific lineages, massive rearrangements, gene duplications, loss or gain have been observed e.g., in Campanulaceae (Knox 2014). These
variations provide sufficient unique attributes to reconstruct phylogenetic relationships with strong statistical support, and to investigate the origin and evolutionary patterns of plastids (Pouchon et al. 2018; Tonti-Filippini et al. 2017) through comparative genomics.

To date, only eight chloroplast genomes have been sequenced and reported from three genera in the Senecioneae tribe (Doorduin et al. 2011; Lee et al. 2016; Wang et al. 2019). In this study, the first five and six chloroplast genomes in Senecio and Dendrosenecio respectively were sequenced and analysed. The objectives were: to generate, characterize and analyse the complete chloroplast genomes of 11 species of Senecioneae; to identify highly variable regions that could be of phylogenetic utility within the tribe through comparative analyses and; to investigate the potential of chloroplast phylogenomics in resolving phylogenetic relationships among the species of Senecioneae, with key interest on the phylogenesis of Senecio and Dendrosenecio.

**Methods**

**Plant material and genome sequencing**

Fresh young leaves of 11 species of Senecioneae were collected from the tropical mountains in eastern Africa (Table 1). The species were identified according to the morphological descriptions given in the Flora of Tropical East Africa (Beentje et al. 2005; Knox 2005). Voucher specimen for each species was deposited in the Herbarium of Wuhan Botanical Garden, Chinese Academy of Sciences (HIB). Total genomic DNA was extracted from approximately 100 mg of leaf material for each sample using a modified 2 × cetyltrimethylammonium bromide (CTAB) method (Doyle 1987). The DNA quality was checked using Qubit 2.0 Fluorometer (Life Technologies, CA, USA). A DNA library was constructed for each species by shearing the genomic DNA into short fragments of ~ 350 bp. The DNA was sequenced based on the pair-end sequencing technique implemented on an Illumina Hiseq 2500™ platform (Illumina Inc., San Diego, CA, USA). An average of 22.75 million paired reads, at least 5 Gb of raw sequence data, were generated for each species.

**Genome assembly and annotation**

The raw data were filtered and trimmed using Fastp software using the default settings (Chen et al. 2018); all low-quality reads were discarded. The de novo assembly of the filtered reads, into complete chloroplast genomes, was performed using NOVOPlasty (Dierckxsens et al. 2017) with default seed and K-mer = 31–39. The contigs were then mapped to the chloroplast genomes of Jacobaea vulgaris Gaertn. (NC_015543; Doorduin et al. 2011) and Pericallis hybrida (Regel) B. Nord. (NC_031898; Wang et al. 2019) using Geneious Prime 2019 (Biomatters Ltd., Auckland, New Zealand; https://www.geneious.com). Basic local alignment search tool ver. 2.2.18+ (Camacho et al. 2009) was used to ascertain the positions of the single copies and the inverted repeat regions by self-blasting the assembled sequences.

GeSeq (Tillich et al. 2017) was used to annotate each of the chloroplast genomes using the complete chloroplast genome sequences of Jacobaea vulgaris and Pericallis hybrida as references. Where necessary, manual corrections were performed in Geneious Prime 2019 (Biomatters Ltd., Auckland, New Zealand), to rectify the start and stop codons of the protein-coding genes (PCGs), based on the annotations of J. vulgaris and P. hybrida. A circular genome map for each species was generated using OGDdraw v1.2 (Lohse et al. 2007). All annotated genome sequences were submitted to the GenBank under the accession numbers listed in Table 1.

**Codon usage and microsatellite repeats identification**

The level of codon usage bias was determined by analysing the Relative Synonymous Codon Usage (RSCU; Sharp and Li 1987), Effective Number of codon (ENC; Wright 1990) and the Codon Biased Index (CBI; Morton 1993) for all PCGs, in DnaSP 6.10 (Rozas et al. 2017). The frequency of amino acid was also considered. The MicroSatellite Identification tool Perl Script (MiSa; Thiel et al. 2003), was used to mine for SSRs with the parameters set at 10 for mononucleotides, 5 for dinucleotides, 4 for trinucleotides and 3 for tetra-, penta- and hexa-nucleotides.

**Genome comparative analyses and divergence hotspot identification**

The available chloroplast genomes of Asteraceae species have been shown to harbour no major differences in their sizes, gene content and arrangement. The whole genome size, GC percentage, LSC, SSC, IR and number of gene in each of the 11 species, were therefore compared to three other species of Senecioneae. Preliminary comparative analyses among the species within each genus revealed highly conserved sequences with > 99% pairwise identity and > 98% identical sites. Consequently, one chloroplast genome sequence was randomly picked per genus to conduct further comparative studies against other chloroplast genomes within the tribe. The expansion/contraction of the IR regions was assessed by comparing the positions of SC/IR junctions and their adjacent genes using IRscope (Amiryousefi et al. 2018).

Further, to outline any significant sequence divergence spots and genome rearrangements, the chloroplast genomes were aligned and plotted in MAUVE (Darling et al. 2004), with Nicotiana tabacum L. (NC_001879; Shinozaki et al. 1986) being used as an
| General genome characteristic | Source | Herbarium accession numbers | GenBank accession number | Genome size (bp) [GC%] | LSC length (bp) | SSC length (bp) | IR length (bp) | PCGs (duplicated in the IR) | Tmas (duplicated in the IR) | Rrna (duplicated in the IR) |
|-------------------------------|--------|----------------------------|--------------------------|-----------------------|---------------|---------------|---------------|--------------------------|--------------------------|--------------------------|
| Dendrosenecio johnstonii (Oliv.) B. Nord. | Mt. Kilimanjaro | SAJIT-002716 | MG560050 | 150,607 [37.5] | 83,469 | 17,754 | 24,692 | 80 (8) | 30 (7) | 4 (4) |
| D. meruensis (Cotton & Blakelock) E.B. Knox | Mt. Meru | SAJIT-002556 | MG560049 | 150,587 [37.5] | 83,448 | 17,755 | 24,692 | 80 (8) | 30 (7) | 4 (4) |
| D. elgonensis subsp. elgonensis (T.C.E.Fr) E.B. Knox | Mt. Elgon | SAJIT-003220 | KY434194 | 150,548 [37.5] | 83,403 | 17,771 | 24,687 | 80 (8) | 30 (7) | 4 (4) |
| D. keniodendron (R.E. Fr. & T.C.E.Fr) B. Nord. | Mt. Kenya | SAJIT-002100 | KY434193 | 150,555 [37.5] | 83,418 | 17,755 | 24,691 | 80 (8) | 30 (7) | 4 (4) |
| D. battiscombei (R.E. Fr. & T.C.E.Fr) E.B. Knox | Mt. Kenya | SAJIT-002802 | KY434195 | 150,556 [37.5] | 83,410 | 17,762 | 24,692 | 80 (8) | 30 (7) | 4 (4) |
| D. brassiciformis (R.E.Fr. & T.C.E.Fr) Mabb. | Aberdare Ranges | SAJIT-003289 | MG560051 | 150,551 [37.5] | 83,424 | 17,747 | 24,690 | 80 (8) | 30 (7) | 4 (4) |
| Senecio moorei R.E.Fr. | Mt. Kenya | SAJIT-201834 | MH483949 | 151,204 [37.2] | 83,278 | 18,300 | 24,813 | 80 (8) | 30 (7) | 4 (4) |
| S. kenophyllum R.E.Fr. | Mt. Kenya | SAJIT-201831 | MH483946 | 151,413 [37.2] | 83,422 | 18,277 | 24,857 | 80 (8) | 30 (7) | 4 (4) |
| S. purshschelleri Engl. | Mt. Kenya | SAJIT-201832 | MH483947 | 151,191 [37.2] | 83,243 | 18,302 | 24,823 | 80 (8) | 30 (7) | 4 (4) |
| S. schweinfurthii O.Hoffm. | Mt. Kenya | SAJIT-201833 | MH483950 | 151,260 [37.2] | 83,255 | 18,389 | 24,823 | 80 (8) | 30 (7) | 4 (4) |
| S. roseiflorus R.E.Fr. | Mt. Kenya | SAJIT-201835 | MH483948 | 151,228 [37.2] | 83,329 | 18,279 | 24,810 | 80 (8) | 30 (7) | 4 (4) |
| Jacobea vulgaris Gaertn. | – | – | NC_015543 | 150,689 [37.3] | 82,855 | 18,277 | 24,777 | 81 (8) | 29 (7) | 4 (4) |
| Ligularia fischeri (Ladeb.) Turcz. | – | – | KT988070 | 151,133 [37.5] | 83,238 | 18,233 | 24,831 | 80 (8) | 29 (7) | 4 (4) |
| Pericallis hybrida (Regel) B.Nord | – | – | NC_031898 | 151,267 [37.3] | 85,751 | 18,331 | 23,591 | 79 (8) | 30 (7) | 4 (4) |
external reference genome. Nucleotide diversity (Pi) in the non-coding regions (> 200 bp) of the five species of Senecioneae was analysed in DnaSP v6.10 (Rozas et al. 2017). Potential primers for ten sites with the highest Pi values were designed using Primer3 (Untergasser et al. 2012) using default settings.

Phylogenetic analyses
A total of 75 species, representing 49 genera from 11 tribes of Asteraceae, were downloaded from the NCBI (Additional file 1: Table S1) for phylogenetic analyses. Besides, data for Adenophora divaricata Franch. & Sav. and A. stricta Miq. (Cheon et al. 2017) were downloaded and used as outgroups in this analysis. Sequences of 70 common PCGs were extracted from the 77 species. Each gene was separately aligned using MUSCLE (Edgar 2004) and then concatenated in Geneious Prime 2019 (Biomatters Ltd., Auckland, New Zealand). Phylogenetic reconstructions were carried out using Maximum Likelihood (ML) and Bayesian Inference (BI) methods. Each method was used twice in independent analyses based on unpartitioned and partitioned data. Before the ML analysis using unpartitioned data, the best-fit DNA substitution model was determined using ModelFinder (Kalyaanamoorthy et al. 2017) as implemented in IQ-TREE version 1.5.4 (Nguyen et al. 2015). Maximum Likelihood (ML) analysis was conducted using IQ-TREE with a bootstrap analysis of 5000 replications under the GTR + F + R6 nucleotide substitution model. MrBayes v3.2.6 (Ronquist et al. 2012) was used to implement the BI analyses based on the unpartitioned data set, using four independent Markov Chain Monte Carlo runs with three heated and one cold chain. The chains were run for $2 \times 10^6$ generations with sampling from the cold chain run after every $10^3$ generations. The analysis was stopped after the average standard deviation of split frequencies as calculated by Mr. Bayes was below 0.01, an indication that convergence had been attained. The first 25% of all generations were excluded, and a consensus phylogenetic tree was obtained based on majority rule from the remaining trees. Branch support was indicated by posterior probability (PP) values. The data set was then partitioned by categorizing the nucleotides in each gene based on the position (first, second, or third) they occupy in a codon. The best partitioning scheme and substitution models were calculated using PartitionFinder2 (Lanfear et al. 2017). The generated phylogenetic trees were visualised and formatted in Interactive tree of life (iTOl) v3 (Letunic and Bork 2016).

Results
Chloroplast genome organization and content
An average of 22.6 million (95.2%) clean reads were generated for each species. The chloroplast genomes of the two genera were comparable in terms of structural organization, gene content, and arrangement. The overall chloroplast genome size varied slightly within each genus, but significantly between the genera ranging within 150 kb in Dendrosenecio and to 151 kb in Senecio. Each of the chloroplast genomes had four regions including a large single copy of ~83.5 kb in both genera, two inverted repeats ~24.7 kb, and a small single copy ~17 kb in Dendrosenecio and ~18 kb in Senecio. The GC percentage values of the entire genome and for each of the sections were identical in all species within a respective genus (Table 1). Each of the plastid genomes encoded a total of 114 unique genes of which 80 were PCGs, 30 tRNAs and four rRNAs. All the PCGs, except three, had the standard AUG start codon (Table 2). Seventeen genes; 11 PCGs and six tRNA genes contained either one or two introns. Eighteen genes were duplicated in the IR regions, with rps12 being uniquely positioned with its 5′ end exon at the LSC and the other is located in the IR regions. Both ycf1 and rps19 genes also had their 3′ ends duplicated on the IR regions (Fig. 1; Table 1).

Codon usage and microsatellite sequences
The total sequence size coding for protein genes was 78,879–79,146 bp in Dendrosenecio and 78,203–78,531 in Senecio. These protein sequences encoded 26,293–26,382 and 26,067–26,177 codons respectively, including stop codons. Leucine was encoded by the highest number of codons, with RSCU values of the entire genome and for each of the sections ranging from 0.288 to 0.315 (Additional file 2: Table S2). The average RSCU and amino acid frequency values for each species were plotted using R-script (Fig. 2; Zhang et al. 2018).

In each genus, the analyses of repetitive sequences revealed minimal variation in the total number and position of each repeat motif, most of which were shared among the chloroplast genome sequences. In Dendrosenecio, 340 microsatellites were discovered while the species of Senecio had 331 SSR repeats. On average, species of Senecio had the highest number of mono-, tri-, tetra- and hexa-nucleotides, while di- and penta-nucleotides were the most abundant repeats.
Table 2 List of genes identified in the studied chloroplast genomes of 11 species of Senecioneae

| Gene family                  | Gene                                                                 |
|------------------------------|----------------------------------------------------------------------|
| Transfer RNA                 | trnA-UGC*, trnM-CAU, trnI-GAU*, trnM-CAU, trnR-ACG, trnS-UGA, trnC-GCA,  |
|                              | trnG-CC, trnK-UCU*, trnM-GUU, trnW-CCA, trnF-GGU, trnD-GUC, trnV-UAC*,  |
|                              | trnL-CAA, trnY-GUA, trnR-UCU, trnT-UCU, trnE-UGG, trnH-GUG, trnL-UAA*,  |
|                              | trnP-UGG, trnS-GCU, trnV-GAC, trnF-GAA, trnI-CAU, trnL-UAG, trnQ-UUG,     |
|                              | trnS-GGA, trnG-UCC*                                                  |
| Small ribosomal units        | rpo2,3,4,7,8,11,12*, 14,15,16*, 18,19,19P                           |
| Large ribosomal units        | rpl2*, 14,16*, 20,22,23,32,33,36                                   |
| RNA polymerase sub-units     | rpoA, rpoB, rpoC1*, rpoC2                                            |
| Translation initiation factor | infA                                                                 |
| NADH dehydrogenase           | ndhA*, ndhB*, ndhC, ndhD*, ndhE, ndhF, ndhG, ndhH, ndhI, ndhJ, ndhK    |
| Photosystem I                | psaA, psaB, psaC, psaI, psaJ, ycf3**, ycf4                           |
| Photosystem II               | psbA, psbB, psbC, psbH, psbI, ycf4                                    |
| Cytochrome b/f complex       | petA, petB*, petD*, petG, petL, petN                                 |
| ATP synthase                 | atpA, atpB, atpE, atpF*, atpH, atpI                                 |
| Large subunit of rubisco     | rbcl                                                                  |
| Maturase                     | matK                                                                  |
| Protease                     | clpP**                                                                |
| Acetyl-CoA-carboxylase sub-unit | accD                                                              |
| Envelope membrane protein    | cemA                                                                  |
| Component of TIC complex     | ycf1                                                                  |
| c-Type cytochrome synthesis  | ycf2, ycf15                                                           |
| Hypothetical genes reading frames | rrn4.5, rrn5, rrn16, rrn23                                         |

* Genes with a single intron, ** genes with two introns, * trans-spliced genes, b genes with an alternative start codon

Fig. 1 A representative chloroplast genome map of a Dendrosenecio and b Senecio. Genes are color-coded based on their function as shown in the legend. The inner circle indicates the inverted repeat boundaries and the genome’s GC content. The arrows indicate the direction of gene transcription.
Certain repeat motifs were genus-specific while a few e.g., C/G and AGCTAT/AGCTAT in *D. johnstonii* and AATCT/AGATT and AATTC/AATTG in *S. keniophytum* were species-specific (Additional file 3: Table S3). The present SSRs were classified based on the variation in repeat type, the number of repeats in each motif, presence or absence of the repeat, and the position of each repeat in the genome. Microsatellites were considered polymorphic if they showed variation, were present in all plastid genomes and were positioned at homologous regions across all species in each genus. Based on this criterion, 25 polymorphic SSRs were discovered in *Senecio* and only five in *Dendrosenecio* (Tables 3, 4).
Genome comparative analyses

In each genus, the newly annotated chloroplast genomes had no significant differences, except for the slight variations in size and gene positioning. The size difference between the largest and the smallest genome among the *Dendrosenecios* and *Senecio* was 59 bp and 222 bp respectively. Based on the currently available chloroplast genomes, *Dendrosenecios* had the smallest chloroplast genome size within the Senecioneae tribe, with a difference of 82 bp from the immediate largest cp genome (*Jacobaea vulgaris*; Table 1). The genes adjacent to the IR/SC junctions (trnH, rps19, ycf1 and ndhF) were similar in all species, except in *Pericallis hybrida* which had rpl2 at the LSC/IR junction. The IRb region expanded into the coding region of rps19, resulting into a pseudogene (ψ) of varying length at the IRa in all but in *P. hybrida* and *S. moorei*. In *S. moorei*, a 14 bp gap was observed between the rps19 gene and the JLB border. At the JSB junction, the IRb expanded into the coding region of ycf1 gene; thus ψycf1 appeared on the IRa region in all species. Two genes, ndhF and ψycf1, were positioned at varying points adjacent to the JSA junction. In *I. vulgaris* and *Ligularia fischeri* (Ladeb.) Turcz., the ndhF and ψycf1 genes overlapped. The JLA junction was uniformly flanked between ψrps19 and trnH genes, except in *S. moorei* and *P. hybrida* where rpl2 and trnH bordered the junction. Figure 4 shows the genes adjacent to the junctions and their order in representative genomes from each genus. In *Senecio* two species were used to show the differences recorded at JLA and JLB in *S. moorei*.

There were no major rearrangements detected among the newly sequenced chloroplast genomes, an indication that chloroplast genomes within these two genera could be much conserved (Fig. 5). However, the existence of two inversions in the LSC region in reference to *Nicotiana tabacum* was identified in all newly generated chloroplast genomes. The arrangement of genes in the SSC region was also different in the Asteraceae species, apart from *L. fischeri*, whose alignment was identical to that of *N. tabacum* (Additional file 4: Figure S1). The nucleotide polymorphism test identified 74 sites with Pi values ranging from 0.00089 (ndhB–ndhB) to 0.06852 (trnH-GUG-psbA). Figure 6 indicates the regions with high levels of intergeneric variation (Pi values > 0.01). Potential PCR primers were designed for the ten highly polymorphic sites (Table 5).

Phylogenetic relationships

The final sequence alignment of common protein-coding genes had 60,992 characters in 70 chloroplast genome loci for 77 taxa. Phylogenetic relationships among the 75 species representing ten tribes of Asteraceae were unveiled based on ML and BI analyses. The tribes were recovered as monophyletic clades each with significant statistical support in all the generated trees. Intergeneric relationships within tribe Senecioneae, to which the 11 newly sequenced chloroplast genomes belong, were clearly defined and strongly supported in all data schemes (100% BS and 1.0 PP). The phylogenetic analyses strongly supported three sub-clades within the tribe Senecioneae; one that comprised of the genus *Ligularia,*
the second sub-clade contained both *Senecio* and *Jaco-baea* while the third one had species from *Dendrosene-cio* and *Pericallis*. The sister relationship between the species of *Senecio* was congruent in all the analyses, differing only in support values at the clade containing *S. moorei* and *S. schweinfurthii* which was highly supported (BS ≥ 92) in ML trees but gained weak support in BI analyses (PP ≤ 0.5). The six species of *Dendrosenecio* were split into two clades, distinctly separating species from Tanzania and species from Kenya (Fig. 7).

### Table 3 Details of 25 potentially polymorphic microsatellite repeats in five species of *Senecio*

| Microsatellites | Senecio schweinfurthii | S. keniohytonym | S. moorei | S. purtschelleri | S. roseiflorus | Location |
|----------------|-------------------------|----------------|-----------|-----------------|----------------|-----------|
| (AAAT)3;(ATT)4 | 1849 1860 | 1881 1892 | 1860 1871 | 1848 1859 | 1881 1892 | Intron |
| (T10)11;(T16n)(TTAGA)3 | 8685 8694 | 9134 9180 | 8697 8706 | 8689 8699 | 8317 8326 | IGS |
| (T10)(A10);(A10).calc(T10) | 9738 9747 | 9720 9729 | 9738 9760 | 9731 9754 | 9695 9704 | IGS |
| (T17);(10);(16) | 10859 10875 | 10857 10866 | 10872 10887 | 10866 10881 | 10852 10861 | IGS |
| (T11);(23);(16) | 12933 12943 | 12919 12941 | 12945 12955 | 12919 12929 | 12925 12940 | IGS |
| (T16);(16);(15) | 16878 16893 | 18219 18228 | 16890 18256 | 16846 16878 | 16875 16884 | Intron |
| (AT6);(A10);(TA6) | 26849 26860 | 26787 26798 | 26861 26872 | 26834 26845 | 26776 26797 | IGS |
| (T10);(gA11);(T10);(10);(A10);(14) | 28219 28240 | 28166 28179 | 28231 28252 | 28204 28225 | 28154 28174 | IGS |
| (ATAA)3;(T10);(CTA)5 | 32007 32021 | 30007 30016 | 32019 32033 | 31992 32006 | 31008 31022 | IGS |

### Table 4 Details of potentially polymorphic microsatellite repeats in six species of *Dendrosenecio*

| Microsatellites | D. keniodendron | D. elgonensis subsp. elgonensis | D. battiscombei | D. meruensis | D. johnstonii | D. brassiciformis | Location |
|----------------|----------------|-------------------------------|----------------|-------------|----------------|----------------|-----------|
| (A13);(14);(16);(TA29) | 42047 42059 | 42044 42057 | 42051 42064 | 42044 42072 | 42053 42068 | 42049 54600 | Intron |
| (T10);(11);(13) | 59146 59155 | 59130 59140 | 59150 59160 | 59179 59189 | 59200 59210 | 59147 59210 | IGS |
| (A10);(11);(12) | 70469 70478 | 70456 70465 | 70474 70483 | 70368 70378 | 70521 70532 | 70473 70532 | IGS |
| (T11)12;15 | 79748 79762 | 79736 79746 | 79740 79754 | 79781 79791 | 79802 79813 | 79802 79813 | IGS |
| (A,AgaaatttttttggA,AgaaatttttttgAT,AgaaatttttttA)1163940 | 80327 80342 | 80311 80327 | 80319 80334 | 80333 80372 | 80355 80393 | 80355 80393 | IGS |
Fig. 4 Comparison of the large single copy, inverted repeats, and small single copy junction positions in six species of Senecioneae (Asteraceae). Genes adjacent to the junctions are shown as blocks of different colours.

Fig. 5 Comparison of sequence arrangement in the chloroplast genomes of 11 species of Senecioneae (Asteraceae). Conserved orthologs are indicated by locally collinear blocks. Similar blocks among the genomes are coded in one colour and joined by a line. The genes above the line are transcribed in a clockwise direction, those below the line are transcribed towards the counter-clockwise direction.
### Table 5 Details of ten primers that target the most divergent regions in the chloroplast genomes of Senecioneae species

| Orientation | Primers | Tm | Target region | Expected product size |
|-------------|---------|----|---------------|----------------------|
| F           | AAATAGGAGGAAGCTGTGAC | 55 | trnH-psbA    | 603                  |
| F           | GCACATGGCTTCTTCAACCTA | 55 |             |                      |
| F           | CACGTACAGTTCTATGGGA | 55 | ndhD-ccsA   | 658                  |
| R           | TTGGGCCTTTTATCTTGGGA | 55 |             |                      |
| F           | TTTGCGAAAAGAGGAAGACA | 55 | trnT-trnL   | 812                  |
| R           | TCCATACCCAGGCTCAATTC | 55 |             |                      |
| F           | TGTGGTAATTGCCGGTGAAGTA | 55 | ndhI-ndhG  | 759                  |
| R           | TACGTAATAAGGAGCTCGG | 55 |             |                      |
| F           | CTTGCTTCTAATCCAAAT  | 55 | trnS-rps4   | 588                  |
| R           | TGACTCTTCAAACTTCCA  | 54 |             |                      |
| F           | TTGACCTTTGAAAACAACCG | 55 | rps18-rpl20 | 546                  |
| F           | ACAGATACAGTTGCTTCTCTT | 55 |             |                      |
| F           | GGAACCTAAAGAAGAGTCA  | 54 | psbM-trnD   | 880                  |
| R           | ATGTCATGCTTGGTCAATTT | 55 |             |                      |
| F           | GCGGATTCATTTGAACCTCA  | 54 | ndhC-trnV   | 959                  |
| R           | AGCCCTAGAGGCTTATAAAA | 55 |             |                      |
| F           | TAACGCTGCCAAATATCCTT | 55 | rpl32-ndhF  | 1194                 |
| R           | AGAAGAAGTCCCAACCTAT  | 55 |             |                      |
| F           | ATGCCAATTCATCATACTTT | 53 | psbZ-trnG   | 617                  |
| R           | ACTACATATGACGGCTAAC  | 54 |             |                      |

*Tm* annealing temperature, *F* forward primer, *R* reverse primer
inter-specific relationship within the Kenyan species differed in the different phylogenetic trees (Additional file 5: Figure S2).

Discussion

Genome structure and content

It is typical of higher plants to possess a chloroplast genome that has a quadripartite structure and a relatively well-conserved gene content and arrangement (Ravi et al. 2008; Wicke et al. 2011; Yurina et al. 2017). However, minor to major variability in chloroplast genome structure has been observed in specific plant lineages (Guisinger et al. 2011). The current upsurge in the amount of chloroplast genome-scale data has played a crucial role in the enhancement of knowledge on the evolution and organization of chloroplast genomes. Both Senecio and Dendrosenecio are conspicuous genera in the tribe Senecioneae, the former for being the largest genus in the tribe and the latter for exhibiting an atypical growth habit. In this study, the first complete chloroplast genome sequences of five Senecio and six Dendrosenecio species were generated. The genome structures were comparable to those of higher plants each exhibiting four compartments; a pair of inverted repeats and a pair of single copies of unequal length. The overall chloroplast genome size of between 150 and 151 kb, was comparable to the majority of chloroplast genomes of species within the Asteraceae family (Salih et al. 2017). The GC content of species within each of the two genera was identical: 37.2% in Senecio and 37.5% in Dendrosenecio (Table 1, Fig. 1). Usually, chloroplast genomes exhibit a high AT-GC content ratio (Ravi et al. 2008), which is a crucial factor in genome organization and stability (Niu et al. 2017).

Chloroplasts have undergone enormous changes since they evolved from cyanobacteria in over a billion years ago (Timmis et al. 2004). One of the notable transformations is the reduction in size, which is primarily accredited to considerable gene transfer from the chloroplast genome to the nuclear genome (Bock and Timmis 2008; Martin et al. 1998). Generally, the chloroplast genomes of higher plants currently encode 70–90 protein-coding genes, which is approximately 2% of the total PCGs found in cyanobacterium Synechocystis (Eckardt 2006). The number of unique genes encoded in the available chloroplast genomes of Asteraceae varies slightly between 110 and 115 (Dempe-wolf et al. 2010; Doorduyn et al. 2011; Kumar et al. 2009; Lu et al. 2016; Walker et al. 2014). The 11 species reported
here encoded an equal number of 114 unique genes, with 80 genes encoding for protein, similar to the chloroplast genomes of *Taraxacum* F.H. Wigg. (Asteraceae; Salih et al. 2017). The exon of some specific genes has been subjected to interruptions by either a single or several introns. In such cases, the entire sequence, containing both the exon and intron(s), is transcribed into a forerunner RNA and later the introns are detached for accurate production of a proper transcript (Eckardt 2007; Plant and Gray 1988). The *rps12* gene is distinctively placed, with its 5’-end being positioned in the LSC region and the 3’-end is duplicated over (> 1.6) or under- (< 0.6) represented. In particular, only cysteine had the least frequency which is congruent to most Asteraceae species including *Nicotiana tabacum* (Hildebrand et al. 1988). Seventeen genes, in each of the studied species, had either a single or several introns (Table 2). Excluding *rps12* gene whose intron is exceptionally large, *trnK*-UUUU had the largest single intron. Two genes, *ycf3* and *clpP*, had two introns within their exons this arrangement has also been observed in other Asteraceae species including *Artemisia annua* L. (Shen et al. 2017). However, in *Ageratina adenophora* (Spreng.) R.M. King. & H. Rob., the *rpoC1* gene has two introns (Nie et al. 2012), which is considered rare among other Asteraceae species.

**Codon usage and repetitive sequences**

Relative synonymous codon usage values of less than one indicate that the codons are less frequent, > 1 represents those that are more frequently used, whereas = 1 shows no bias (Uddin 2017). An identical trend in the manner in which the amino acids were encoded was discovered among the 11 species (Fig. 2, Additional file 2: Table S2). The usage frequency of leucine was higher than the rest, among the 11 species and except for five, the rest lacked any variations in terms of length and motif. On the contrary, 25 microsatellites in *Senecio* exhibited slight variations based on the same criteria. Being considered polymorphic, the identified microsatellites are therefore potential molecular markers for use in further studies within the respective genus.

**Chloroplast genome comparison**

There were no remarkable structural rearrangements among the taxa of these two genera. The chloroplast genomes are highly conserved with an identical structure and an equal number of genes, an indication that this could be the case in chloroplast genomes of most species of these genera. Comparative analyses against representatives of three other genera of Senecioeae revealed a similar trend of the well-preserved structure and organization. The inverted repeat region is present in a majority of angiosperms chloroplast genomes. Initially the IR was reported to serve as a whole-genome stabilizer by reducing recombination between the two SC regions; however, these reports lacked support as more chloroplast genomes revealed significant rearrangements even with both copies of the IR present (Jansen and Ruhlman 2012). Comparative analyses between plants of different plant lineages revealed that inverted repeats could contract/expand up to a few hundred base pairs even among closely related species (Goulding et al. 1996). In this study, the comparison of the IR/SC junctions showed a slight expansion of the IR in all except in *Senecio moorei* and *Pericallis hybrida*. In the novel chloroplast genomes, the same genes were found adjacent to the junctions, and only slight length variations were recorded in *P. hybrida* and *S. moorei*. In *P. hybrida* the LSC/IRb junction contracted into the *rpl2* gene, whereas it extended into the *rps19* in all the other analysed species. Two pseudogenes (ψ*rps19* and ψ*ycf1*) of varying length were generated in the IRa region, as a result of the expansion of the IR
into the exons of *rps19* and *ycf1* genes respectively (Fig. 4). This pattern of expansion of the IR, and the introduction of partial copies of genes with non-coding abilities represents a familiar phenomenon in majority of Asteraceae species (Wang et al. 2015), and besides being a source of DNA barcodes, it can offer insights into the evolutionary processes of plastid genomes.

The entire sequences of the 11 chloroplast genomes generated here, lack any striking inversions or rearrangements and therefore were outlined as a single locally collinear block in our analyses. However, certain regions harboured divergent sites the majority of which were in the non-coding regions. Among the few sites with significant deviations include *trnH-(UIG)-psbA, ndhD-ccSA, trnT-(UGU)-trnL(UAA), ndhI-ndhG*, and *trnL-UAG-rpl32*. Other regions including *trnL(UAG)-rpl32* and the exons of *ndhF* and *ycf1* were within a conserved block, but they had significant divergent points. These findings were supported by results obtained from the DNA polymorphism test based on genus representatives, as same regions were noted to have high nucleotide variability (Pi). Some of these regions have previously been reported in chloroplast genomes of other species (Salih et al. 2017; Wu et al. 2018) and used in phylogenetic studies of numerous taxa including *Senecio* (Kandziora et al. 2016). Non-coding regions in chloroplast genomes have shown high potential for use as molecular markers for phylogenetic studies at low taxonomic levels in Angiosperms (Shaw et al. 2005). Therefore, the regions identified herein are prospective sources of highly informative markers for phylogenetic utility in elucidating intergeneric relationships within the tribe. Subsequently, ten potential markers were developed, allowing for specific amplification of each of the ten most polymorphic sites.

We compared the Senecioneae chloroplast genomes against *Nicotiana tabacum* and detected the two inversions reported to be shared by all clades of Asteraceae family, apart from species of the Barnadesioideae subfamily (Kim et al. 2005). Six conserved gene blocks were identified among the chloroplast genomes indicating the most conserved regions of the genomes. The SSC region in *Ligularia fischeri* was differently oriented in relation to the other Asteraceae species. This re-inversion is considered an ordinary phenomenon among chloroplast genomes of higher plants, and it is not a product of any evolutionary event, as single-copy regions exist in two equimolar states (Palmer 1983; Walker et al. 2015).

**Phylogenomics analyses**

The rapid increase in the amount of complete chloroplast genome sequences during the past decade, provided essential data to elucidate further and resolve phylogenetic relationships among species. Consequently, in a move towards chloroplast phylogenomics, clarification of phylogenetic relationships at higher and lower taxonomic levels have been achieved (Lu et al. 2015; Ma et al. 2014; Wu et al. 2013). The tribe Senecioneae is often subdivided into three subtribes; Senecioninae, Tussilagininae, and Blennospermataeae (Chen et al. 2011). In this study, the multi-gene analysis resulted in a phylogenetic tree whose branches were strongly supported. The five genera of Senecioneae formed two distinct clades that corresponded to two of the three subtribes of Senecioneae including Senecioninae (*Dendrosenecio, Senecio, Jacobaea* and *Pericallis*) and Tussilagininae (*Ligularia*; Fig. 7, Additional file 5: Figure S2). The 11 newly generated species were well placed within the Senecioneae tribe by both ML and BI phylogenetic methods under partitioned and unpartitioned data schemes. The alpine species of *Senecio* are classified in at least five clades of *Senecio* (Kandziora et al. 2016). The present phylogenetic study recovered a monophyletic group with two sub-clades which split *S. kenophyllum* and *S. roseiflorus* from *S. purshcelleri, S. moorei* and *S. schweinfurthii*. The relationships among the species of *Senecio* was identical in all data schemes differing only in support of the *S. moorei* and *S. schweinfurthii* relationship (94% and 0.3 PP; Fig. 7). The genus *Jacobaea*, just like *Dendrosenecio*, was previously classified in *Senecio* under section *Jacobaea* (Pelser et al. 2002) but later segregated from *Senecio* based on new insights from molecular phylogeny (Pelser et al. 2006). A sister relationship between *Senecio* and *Jacobaea* was highlighted in this study.

Previously, genetic relationship within *Dendrosenecio* species was shown to be strongly correlated to geographic distance (Knox and Palmer 1995) but later segregated from *Senecio* based on new insights from molecular phylogeny (Pelser et al. 2006). A sister relationship between *Senecio* and *Jacobaea* was highlighted in this study.
of *Diplostephium* Kunth (Vargas et al. 2017). Therefore, this study strongly underscored the potential of chloroplast genome-scale data in outlining both inter- and intra-generic phylogenetic relationships within the tribe Senecioneae species and in the family at large. Nonetheless, interspecific relationships were weakly supported and therefore, further comprehensive studies that include more taxon sampling are necessary to enhance our understanding of the evolutionary histories of both *Senecio* and *Dendrosenecio*.

**Conclusion**

*Dendrosenecio* is a segregate genus to *Senecio*. Despite exhibiting some striking morphologic similarities, a few differences existed based on which the two genera were separated. Initially, controversies arose over the segregation although a consensus was later arrived at. Amplified fragment length polymorphism data distinctly separated the two genera, affirming the earlier decisions. However, lack of or limited molecular resources have impeded further studies on the respective genera. This study generated the first complete chloroplast genome sequences in each genus. Chloroplast genomes in both genera are highly similar in structure, gene composition and synteny, but they significantly differ in size. A chloroplast genome multi-gene dataset revealed three strongly supported clades within the tribe Senecioneae, markedly splitting *Dendrosenecio* from *Senecio*. Ten primers, targeting the ten highly divergent regions in the chloroplast genomes of Senecioneae species, were designed. Also, 25 polymorphic cpSSR in *Senecio* and five in *Dendrosenecio* were identified. The ten divergent hotspots could offer the much-needed DNA barcodes for species identification and phylogenetic reconstructions within the tribe, while the cpSSRs provides potential markers for future population-level research in each respective genus.

**Abbreviations**

ITS: internal transcribed spacer; IR: inverted repeat; SSC: small single copy; LSC: large single copy; SSR: simple sequence repeat; CTAB: cetyltrimethylammonium bromide; PCGs: protein-coding genes; RNA: ribonucleic acid; RSCU: relative synonymous codon usage; CUB: codon usage bias; ENC: effective number of codons; ML: maximum likelihood; BS: bootstrap support; BI: Bayesian inference; PP: posterior probability; cpSSRs: chloroplast microsatellites.

**Acknowledgements**

We thank Elizabeth M. Kamande for her support in fieldwork, Josphat K. Saina for his assistance in data analyses, and Justus M. Mulinge for proofreading the manuscript.

**Authors’ contributions**

QF, GH, and JC conceived and designed the experiment. AWG, SA, GH, and JC conducted fieldwork. AWG and SA performed the experiments. AG, SA, and ZL analysed the data. AG wrote the manuscript. ZL assisted in revising the manuscript. All authors read and approved the final manuscript.

**Funding**

This work was funded by Sino Africa Joint Research Center (Nos. Y323771W07 and SAJC201322) and the CAS-TWAS President’s Fellowship Program for developing countries.

**Availability of data and materials**

The datasets generated during the current study are available in the GenBank repository under the accession numbers KY434193–KY434195, MG560049–MG560051 and MH483946–MH483950. All the datasets used for phylogenetic and comparative analyses were downloaded from GenBank, and the accession numbers are provided in the additional files.

**Ethics approval and consent to participate**

Not applicable.

**Consent for publication**

Not applicable.

**Competing interests**

The authors declare that they have no competing interests.

**Author details**

1. Key Laboratory of Aquatic Botany and Watershed Ecology, Wuhan Botanical Garden, Chinese Academy of Sciences, Wuhan 430074, China. 2. University of Chinese Academy of Sciences, Beijing 100049, China. 3. Sino-Africa Joint Research Center, Chinese Academy of Sciences, Wuhan 430074, China. 4. Key Laboratory of Plant Germplasm Enhancement and Specialty Agriculture, Wuhan Botanical Garden, Chinese Academy of Sciences, Wuhan 430074, China.

**Received**: 8 May 2019 **Accepted**: 31 July 2019

**Published online**: 22 August 2019

**References**

Aminoussou A, Hyvönen J, Poczai P (2018) IRscope: an online program to visualize the junction sites of chloroplast genomes. Bioinformatics. https://doi.org/10.1093/bioinformatics/bty220

Beentje HJ, Jeffrey C, Hind DJN (2005) Compositae (Part 3). In: Beentje HJ, Ghazanfar SA (eds) Flora of tropical east Africa. Royal Botanic Gardens, KEW, London, pp 547–869

Bock R, Timmis JN (2008) Reconstructing evolution: gene transfer from plastids to the nucleus. BioEssays 30:556–566. https://doi.org/10.1002/bies.20761

Camacho C, Coulouris G, Avagyan V, Ma N, Papadopoulos J, Bealer K, Madden TL (2009) BLAST+: architecture and applications. BMC Bioinform 10:421. https://doi.org/10.1186/1471-2105-10-421

Chen YL et al (2011) Senecioneae. In: Wu ZY, Raven PH, Hong DY (eds) Flora of China, vol 20–21. Missouri Botanical Garden Press, Beijing, pp 371–544

Chen S, Zhou Y, Chen Y, Gu J (2018) fastp: an ultra-fast all-in-one FASTQ preprocessor. Bioinformatics 34:884–889. https://doi.org/10.1093/bioinformatics/bty560
