The Complete Chloroplast Genome of 17 Individuals of Pest Species *Jacobaea vulgaris*: SNPs, Microsatellites and Barcoding Markers for Population and Phylogenetic Studies

Leonie Doorduin [1,*], Barbara Gravendeel [1,2], Youri Lammers [1], Yavuz Ariyurek [3], Thomas Chin-A-Woeng [4], and Klaas Vrieling [1]

Institute of Biology, Section Plant Ecology and Phytochemistry, Leiden University, PO Box 9505, 2300 RA Leiden, The Netherlands[1]; NCB Naturalis - NHN Leiden University, Einsteinweg 2, 2333 CC Leiden, The Netherlands[2]; Leiden Genome Technology Center, Human and Clinical Genetics, Leiden University Medical Center, PO box 9600, 2300 RC Leiden, The Netherlands[3] and ServiceXS, Plesmanlaan 1d, 2333 BZ Leiden, The Netherlands[4]

*To whom correspondence should be addressed. Tel. +31 (0) 715275117. Fax. +31 (0) 715274900. E-mail: l.j.doorduin@biology.leidenuniv.nl

Abstract

Invasive individuals from the pest species *Jacobaea vulgaris* show different allocation patterns in defence and growth compared with native individuals. To examine if these changes are caused by fast evolution, it is necessary to identify native source populations and compare these with invasive populations. For this purpose, we are in need of intraspecific polymorphic markers. We therefore sequenced the complete chloroplast genomes of 12 native and 5 invasive individuals of *J. vulgaris* with next generation sequencing and discovered single-nucleotide polymorphisms (SNPs) and microsatellites. This is the first study in which the chloroplast genome of that many individuals within a single species was sequenced. Thirty-two SNPs and 34 microsatellite regions were found. For none of the individuals, differences were found between the inverted repeats. Furthermore, being the first chloroplast genome sequenced in the Senecioneae clade, we compared it with four other members of the Asteraceae family to identify new regions for phylogentic inference within this clade and also within the Asteraceae family. Five markers (*ndhC-trnV*, *ndhC-atpE*, *rps18-rpl20*, *clpP* and *psbM-trnD*) contained parsimony-informative characters higher than 2%. Finally, we compared two procedures of preparing chloroplast DNA for next generation sequencing.

Key words: phylogenetic markers; *Jacobaea vulgaris*; Asteraceae; cpDNA polymorphisms; inverted repeat

1. Introduction

Comprising one-tenth of all flowering plants and containing over 20 000 species, the Asteraceae are one of the largest vascular plant families.[1] With the exception of Antarctica, the Asteraceae are distributed on all continents. Species in this family are extremely various in secondary chemistry,[2] inflorescence morphology[3] and chromosome numbers.[4] This huge variation provides great opportunities to acquire insight into the diversification process in this family, which began 42−36 million years ago.[5] The Asteraceae are not only interesting because of their phenotypic and species diversity, but this family also includes members of economically important food crops, herbal species, ornamentals and plants for the cut-flower industry. Other members such as *Jacobaea vulgaris*, *Senecio vulgaris* and *Taraxacum officinale* are weedy and have an economical and ecological impact.[6,7]

We sequenced the complete chloroplast genome of *J. vulgaris* with next generation sequencing

techniques to find new genetic markers that are phylogenetically informative and to discover intraspecific polymorphic markers for population studies. The conservative structure of the chloroplast genome makes it easy to compare with other members of the Asteraceae family. In a recent study of Panero and Funk,[8] 12 major lineages of Asteraceae were found with the Bayesian and maximum parsimony methods by combining 10 chloroplast loci from 108 taxa. Within the subfamily Asteroideae, strong statistical support was found for tribal relationships except for the Senecioneae tribe. In the Bayesian analysis, this tribe was unresolved, and in the maximum parsimony analysis, it was placed as a sister group to Calenduleae without strong statistical support (52% bootstrap proportions). In other studies of Pelser et al.,[9,10] a phylogenetic analysis of the nuclear ribosomal (nr) internal spacers and external spacer and five chloroplast loci were done to clarify intergeneric relationships within Senecioneae and to delimitate the genus *Senecio*. Although these phylogenies gave more insight, they still lacked strong statistical support and resolution.

No chloroplast genome has been previously sequenced from any species in the Senecioneae clade, and the chloroplast genome sequence of *J. vulgaris* can yield more information about variation within this clade, as well as between clades of the Asteroideae subfamily. In this study, the chloroplast genome of *J. vulgaris* (tribe Jacobaea) was compared with *Guizotia abyssinica*, *Helianthus annuus*, *Parthenium argentatum* (all belonging to tribe Heliantheae) and *Lactuca sativa* (tribe Lactuceae). To guide future phylogenetic studies within the Asteraceae family, we identified new phylogenetically informative chloroplast markers by finding differences within and between genome organization.

Jacobaea vulgaris is a troublesome weed that belongs to the Asteraceae family and is native to Europe and western Asia, ranging from Norway through Turkey, and from Great Britain to Siberia. It was first reported in the 1850s in Canada,[11] in 1875 in New Zealand[12] and shortly thereafter in Australia[13] and in 1900 at the west coast of North America.[14] In introduced areas, *J. vulgaris* is a pest species, outcompeting local plants and containing pyrrolizidine alkaloids which are toxic to herbivores.[15] Control is difficult, since the lifecycle can vary from annual to short-lived perennial, depending on the genotype. Moreover, seeds remain viable in the soil for several years.[16] *Jacobaea vulgaris* causes four million dollar losses annually to cattle poisoning and control in Australia alone.[17]

Joshi and Vrieling[18] compared *J. vulgaris* plants from the invasive areas with plants from the native area and found that invasive individuals contained higher pyrrolizidine alkaloid levels, have a 30% higher reproductive effort, are more susceptible to attack by specialist herbivores and less susceptible to generalist herbivores. These results suggest that selection pressures in the invasive area shaped the different allocation patterns in *J. vulgaris* in the invasive areas within 70 generations. However, it is possible that introduced populations were derived from native European populations that happened to express pyrrolizidine alkaloid and allocation patterns that are similar to those currently observed in invasive ranges.

To exclude the null hypothesis that these patterns are observed as a result of genetic drift rather than natural selection, native source populations need to be identified and compared with invasive populations.[19] Source populations can be pinpointed by using neutral molecular markers such as amplification fragment length polymorphisms (AFLPs). A previous study on *J. vulgaris*, based on nuclear AFLP data, did not show a difference in the amount of variation between native and invasive individuals. These findings suggest that introductions from multiple source populations have occurred.[20] Other neutral markers are single-nucleotide polymorphisms (SNPs) and microsatellite markers in the chloroplast genome.[21−23]

Next generation sequencing can produce DNA sequences cheaply and quickly,[24] facilitating the rapid sequencing of nuclear and organellar genomes. Chloroplast genomes are known for their conservative rates of evolution.[25] With an average size of 150 kb, chloroplast genomes are sufficiently large to find differences between and within species.[26−28] The absence of recombination and maternal transmission of the chloroplast genome (limiting gene flow to seed dispersal only) makes cpDNA markers useful for tracing source population(s).[29,30]

In this study, we sequenced the chloroplast genome of 17 *J. vulgaris* individuals by using the Illumina genome analyzer platform. This is the first study sequencing multiple individuals of the same species with next generation sequencing. Multiple individuals were sequenced to reveal intraspecific variation (SNPs and microsatellite loci). Finally, we compared two different procedures of preparation for sequencing the chloroplast genome, namely direct extraction of the chloroplast DNA and amplifying the cpDNA with a long-range PCR.

2. Materials and methods

2.1. Extraction of chloroplasts and isolation of DNA from chloroplasts

By using the chloroplast extraction kit of Sigma-Aldrich (CP-ISO) and following the manufacturer's protocol, chloroplasts from sample nr. 17 (Table 1)

Table 1. Geographical information, percentage of the chloroplast genome sequenced, method used for preparing the template for Illumina sequencing, lane number on the Illumina platform and reads obtained from the 17 individuals of *J. vulgaris* that were sequenced

Sample	Country	Location	Latitude/longitude	% cp genome sequenced	Template sequencing	Illumina lane
1	New Zealand	Haast (South Island)	43°S 169°E	89.9	Long-range PCR	2 (776)
2	Ireland	Caherdaniel	51°N 10°W	88.5	Long-range PCR	2 (545)
3	Norway	Malvik	63°N 10°E	83.4	Long-range PCR	2 (543)
4	Canada	Cardigan	46°N 62°W	89.8	Long-range PCR	2 (838)
5	UK	Padstow	50°N 4°W	98.3	Long-range PCR	2 (1043)
6	Poland	Warsaw	52°N 18°E	94.3	Long-range PCR	2 (650)
7	Spain	Covadonga	43°N 04°W	91.5	Long-range PCR	2 (457)
8	France	Perrogney	47°N 05°E	89.9	Long-range PCR	2 (558)
9	Hungary	Lénárddaróc	48°N 20°E	86.7	Long-range PCR	2 (80)
10	The Netherlands	Ameland	53°N 05°E	88.6	Long-range PCR	2 (468)
11	Australia	Barramonga	38°S 143°E	90.6	Long-range PCR	2 (680)
12	Australia	Franklin (Tasmania)	43°S 147°E	91.8	Long-range PCR	2 (465)
13	UK	Portsmouth	50°N 01°W	98.9	Long-range PCR	2 (1102)
14	Sweden	Kapellskär	59°N 53°E	99.9	Long-range PCR	3 (11 084)
15	New Zealand	Opunake (North Island)	39°S 173°E	94.7	Long-range PCR	2 (691)
16	Germany	Halle	51°N 11°E	98.7	Long-range PCR	2 (805)
17	Spain	Covadonga	43°N 04°W	99.9	Chloroplast DNA extract	1 (18 646)[a]

Numbers given in parenthesis are the number of single-end reads × 1000.
[a]Paired-end reads.

were isolated out of 30 g of fresh leaf material. To remove unwanted whole cells and cell wall debris, the blended leaf material with the chloroplast isolation buffer was centrifuged. To separate the intact from the broken chloroplasts, a 40% Percoll layer was used. Before DNA extraction, the intact chloroplasts were treated with ST buffer (400 mM sucrose, 50 mM Tris, pH 7.8, 0.1% bovine serum albumin) with a final concentration of 25 μg/ml DNAse-1 (Sigma-Aldrich) per gram of leaf material to digest DNA outside the intact chloroplasts. After centrifuging, the chloroplast pellet was resuspended in a Tris EDTA NaCl buffer (100 mM Tris, pH 7.2, 50 mM ethylenediaminetetraacetic acid (EDTA), 100 mM NaCl, 0.2% β-mercaptoethanol). To extract the DNA from the chloroplasts, the chloroplasts were lysed with 1% sodium dodecyl sulphate followed by a phenol/chloroform step to remove proteins. The DNA was precipitated overnight with 1/10 vol. of 5 M ammonium acetate and 1 vol. of isopropanol. After centrifuging, the pellet was washed with 70% ethanol and redissolved in TE buffer (1 M Tris−HCl, pH = 8.0, 0.5 M EDTA).[31]

2.2. Total DNA extraction

Total DNA extractions from samples 1 to 16 of *J. vulgaris* (Table 1) were carried out on five leaf punches of 1 cm diameter each, using the CTAB extraction protocol of Doyle and Doyle.[32]

2.3. Long-range PCR

To develop primers for a long-range PCR, the sequences of *H. annuus* (NC007977), *L. sativa* (DQ383816) and *G. abyssinica* (EU549769) were aligned with BioEdit. With the aid of this alignment and the annotation of *H. annuus*, primers were designed in conserved regions of genes. A total of 18 primer pairs was designed by Primer3 software,[33] which collectively amplified the total chloroplast genome of *J. vulgaris* with overlapping fragments resulting in amplicons between 5808 and 11 110 bp (see Supplementary Table S1 for primer sequences). For amplification, the Takara La Taq kit (Takara Bio Inc., Otsu, Shiga, Japan) was used. PCR was carried out in a total volume of 20 μl containing 8−80 ng of DNA, 2.5 mM MgCl$_2$, 2.5 mM of each dNTP, 0.7 μM of each primer and 1 U *Taq* DNA polymerase. The PCR cycling conditions were as follows: 1 min at 94°C; 30 cycles of 10 s at 98°C and 12 min at 69°C; followed by 10 min at 72°C. PCR products were loaded on a 1.5% agarose gel, stained with ethidium bromide and visualized under UV light to check for amplification. If the PCR products contained more than one band, the total product was always loaded on a 1% agarose gel and bands of the right size were cut out of the gel. To extract and purify the DNA fragments from the gel, the Wizard SV gel and PCR Clean-Up System of Promega was used. All

cleaned PCR products were run on a gel to estimate the amount of product, and in addition, the amount of DNA was quantified with an ND-1000 spectrophotometer (Nanodrop Technologies). All 18 amplicons for each individual sample were pooled in equal molar ratios containing roughly 200–300 ng of DNA resulting in 16 pooled samples of 75 µl each.

2.4. Sequencing

For sequencing of the cpDNA, three lanes on an Illumina sequencer (Illumina 1G/Solexa, Illumina Inc., San Diego, CA, USA) were used. Sequencing was carried out at the Leiden Genome Technology Center. In the first lane, the DNA isolated from the chloroplasts of sample 17 was run with paired-end reads of 32 bp. In the second lane, the pooled long-range PCR products of samples 1–13, 15 and 16 were run and in the third lane sample 14. Both were single-end runs of 35 bp (Table 1). Sample 14 was run in a separate lane because of its low DNA concentration. Preparation of all products was done following the protocol of Illumina kits with minor modifications. For sample 17, DNA was fragmented by a nebulizer using 32 psi N_2 for 6 min. After purification, the DNA was eluted in 15 µl elution buffer. The samples were blunt-ended with T4 DNA polymerase, Klenow polymerase and T4 polynucleotide kinase. After purification, an A-residue was added to the 3′ end of the DNA fragments using Klenow fragment (3′–5′ exo-minus). Purification was done with a Qiagen MinElute column. Adapters of the paired-end adapter oligo mix were ligated to the DNA fragments. After purification with a Qiagen MinElute column, adapter-ligated DNAs in the range of 200–250 bp were size selected using agarose electrophoresis. Products were isolated from the gel using a QIAquick Gel Extraction Kit and after purification a PCR was done.

For samples 1–16 (Table 1), sonication with a Bioruptor was used to fragment the DNA. This machine was placed in a room at 4°C and was kept cool by adding ice. For a total of 15 min, the machine was set on 30 s active and 30 s inactive. This sonication step was repeated four times. All other steps were the same as done for sample 17 except for the PCR step. Unique index tags of six bases provided in the Multiplexing Sample Preparation Oligonucleotide Kit were added in the PCR step to discriminate between the 16 samples. The amplified libraries were quantified by lab-on-a-chip (Agilent Technologies) followed by equimolar mixing of 10 nM per sample. Cluster generation was performed after applying 6 pM of each sample to the individual lanes of the Illumina flow cell, and sequencing was carried out on the Illumina Genome Analyzer according to the manufacturer's instructions.

Image analysis and base calling were performed using the Illumina Pipeline 1.3.2, where sequence tags were obtained after purity filtering. This was followed by an alignment using MAQ.

2.5. Data filtering and genome assembly

Sample 17 from the first lane was used to assemble a draft chloroplast (cp) genome of *J. vulgaris*. The software package MAQ v0.5.0 was used to map all quality-filtered paired reads of the first run against the chloroplast genome of *H. annuus*. To solve gaps in this consensus sequence, a *de novo* assembly was done with the same data using the software package Velvet v 0.6[34] (parameters: hash length = 21), which produced 37 747 contigs. To find contigs with homology to the reference, these contigs were aligned to the *H. annuus* reference sequence with the program Mummer v3.0.[35] The contigs having homology to the reference were extended by using the original reads with Velvet. These extended contigs were aligned to the reference of *H. annuus* with Mummer once again, and the contigs which assembled properly were saved. These final contigs were aligned against the consensus sequence; as a result, some of the gaps in the consensus were solved. A new MAQ alignment was performed, mapping all the Illumina reads against the last consensus sequence made, to produce the draft sequence.

### 2.6. Bridging the gaps that were still in the draft sequence

The draft sequence still contained 23 gaps with an average gap length of 394 bp. Gaps were bridged by adding the data from the runs of the cpDNA amplified by a long-range PCR of 16 individuals. These data were used in Velvet to produce a *de novo* sequence (parameters: hash length = 21, short-fastq reads). The resulting *de novo* contigs were aligned against the draft sequence in the Blast's bl2seq multiple sequence aligner. In this way, five gaps with a total of 1822 bp were bridged. The last 18 gaps were bridged by developing primers around the gaps, and traditional Sanger sequencing to yield the final complete cp genome.

2.7. Annotation

The program DOGMA[36] was used for annotating all genes and to identify rRNAs and tRNAs. A circular cp genome map (Fig. 1) was drawn using the program GenomeVx.[37]

2.8. Comparison of the chloroplast DNA of J. vulgaris with other Asteraceae genomes analysed

A total of 22 conserved protein-coding genes from five species, extracted from all available complete
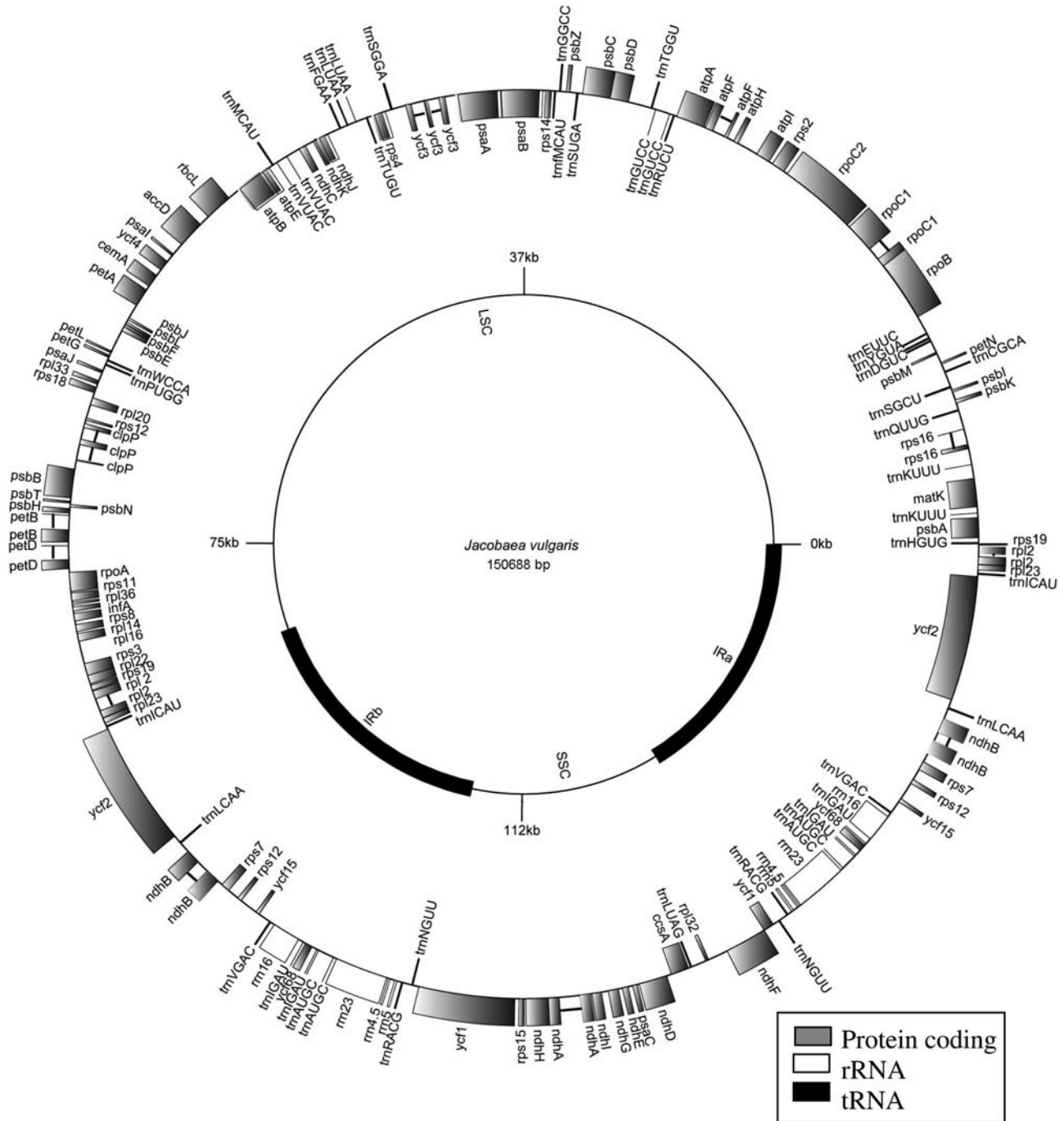
Figure 1. Representative map of the chloroplast genome of *J. vulgaris* (GenBank accession HQ234669).

chloroplast genomes from Asteraceae deposited at NCBI GenBank (*H. annuus*, NC007977; *L. sativa*, DQ383816; *P. argentatum*, GU120098; *G. abyssinica*, EU549769 and *J. vulgaris*, HQ234669), were aligned using the pairwise automatic alignment tool in MacClade 4.06[38] with further adjustment by hand. To get insight in the informative character of the selected protein-coding genes, maximum parsimony analyses were run on the individual alignments comprising a total of 33 669 bp with PAUP* 4.0b10[39]

using heuristic search, random addition with 100 replicates and tree bisection-reconnection (TBR) swapping. The relative robustness for clades found in all single most parsimonious trees (MPTs) was assessed by performing 1000 replicates of bootstrapping[40] using fast, stepwise additions, TBR branch-swapping with 10 random taxon additions per replicate, MULTREES on and holding 100 trees per replicate. We also calculated tree lengths and consistency index (CI) and retention index (RI) values measuring the extent of homoplasy.

2.9. Detection of polymorphic loci

For visualizing the output of all reads, Mapview was used.[41] This program visualizes all reads that are mapped against the reference genome. Furthermore, it can produce a SNP list. The final assembled cp genome was used as a reference. To find SNPs, genomes of individuals 1–17 were used. SNPs were only added to the list if at least one individual that varied from the reference genome had a coverage of at least 30 reads traversing that particular nucleotide and only when SNPs were located outside A and T polymer regions. Potential microsatellite regions were tracked by looking for 10 or more repeats of A and T nucleotides.

3.   Results and discussion

3.1.   Construction of the chloroplast genome of *J. vulgaris*

The chloroplast genome of *J. vulgaris* is 150 686 bp in length. The genome contains two inverted repeat (IR) regions of 24 777 bp each. The IRs are separated by a large single-copy (LSC) and a small single-copy (SSC) region of 82 855 and 18 277 bp, respectively. The genome comprises 81 protein-coding genes of which seven are located in the IRs. *Ycf1* lies partly in the IR and the single-copy region. The four rRNA genes are all located in the IR. There are 29 unique tRNA genes. Twenty-two tRNA genes are located in the single-copy region, whereas the others are located in the IR (Fig. 1).

The single lane on Illumina yielded sufficient reads to map more than 99.9% of the complete cp genome of *J. vulgaris*. For the pooled individuals, on average 92% of the whole genome was mapped. There was a highly significant correlation between the number of reads and percentage of the genome mapped (Fig. 2). From the figure, it is estimated that ~1 300 000 single-end Illumina reads of 32 bp are needed to reach a mapping percentage higher than 99.9% of the cp genome of *J. vulgaris*.

3.2.   Comparison of the sequencing success of cpDNA extracted from chloroplasts with amplified cpDNA using long-range PCR

For the first lane with cpDNA extracted from isolated chloroplasts, a paired-end run was carried out on the Illumina platform, yielding 582 Mb of sequence with a read length of 32 bp. Of all reads, only 2.1% (391 604 reads) mapped against the chloroplast genome of *H. annuus*. The obtained reads covered 99.9% of the cp genome of *J. vulgaris* (Table 1, Fig. 3). The average coverage was 83 with a coefficient of variation of 0.34 (Fig. 4).
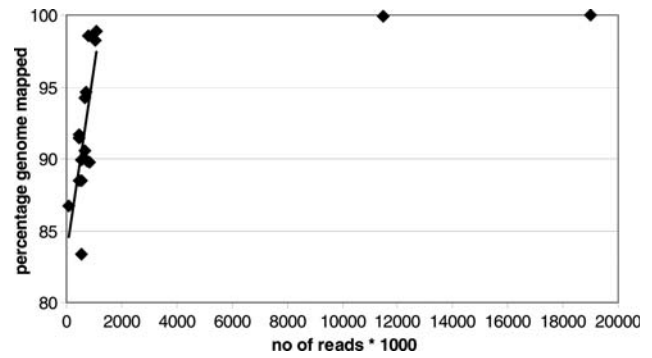

Figure 2. Number of Illumina sequencing reads plotted against percentage of the chloroplast genome mapped for 17 individuals of *J. vulgaris*.

For the other two lanes, containing long-range PCR products of 15 individuals in one lane and the long-range PCR products of one individual in a separate lane, a single-end run was carried out on the Illumina platform. This run yielded reads of 35 bp resulting in 339 and 388 Mb of sequence, respectively. For both lanes, more than 99.9% of the reads (96 894 177 and 11 075 400, respectively) mapped against the chloroplast genome of *H. annuus*. In both lanes, the reads obtained covered more than 99.9% of the cp genome of *J. vulgaris* (Table 1, Fig. 3). The average coverage obtained for both lanes combined was $4920\times$. Average coverage varied largely between primer pairs, with average coverage ranging from $542\times$ for the lowest to $19\,755\times$ for the highest primer pair (Fig. 4). The average coefficient of variation of coverage within primer pairs, averaged over all primer pairs, is 1.04 (Fig. 4). In summary, the variation was three times higher than that obtained with the direct cpDNA extraction.

However, extraction of chloroplasts and subsequent extraction of DNA from these chloroplasts were not very efficient for sequencing the complete chloroplast genome. The cpDNA extract still contained around 98% of non-cpDNA. The low efficiency of the chloroplast extraction method might be due to the fact that (nuclear) DNA sticks to the surface of the chloroplast[31] or to a shortage of DNAases to remove DNA in the intact chloroplast solution. Furthermore, the low efficiency can be caused by poor lysis of the chloroplasts. In contrast, the cpDNA amplified with a long-range PCR contained less than 1% non-cpDNA. Apparently, the long-range PCR worked very efficiently in *J. vulgaris*, and the results were much better than the results obtained with the same method for *Pinus* cpDNA sequencing[42] where non-cpDNA ranged from 19% to 24%.

Although the number of cpDNA reads obtained with the chloroplast extraction method was far lower than that obtained with the long-range PCR methods, the variation in coverage over the total

chloroplast genome was approximately three times lower (Fig. 4). Moreover, the variation in the coverage of the long-range PCR products was primer-dependent (Fig. 4B). Despite the higher variation in coverage, using the long-range PCR products as templates for Illumina sequencing was far more efficient than using cpDNA directly. Moreover, the cpDNA extraction method proved to be cumbersome because we needed 30 g of fresh material per individual.
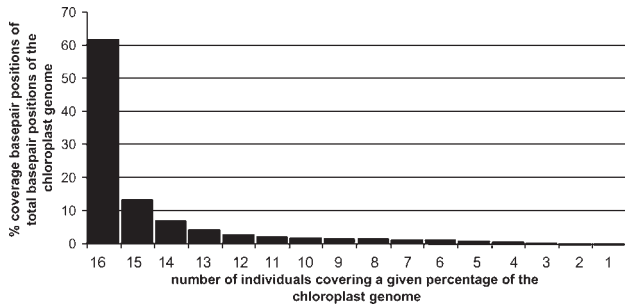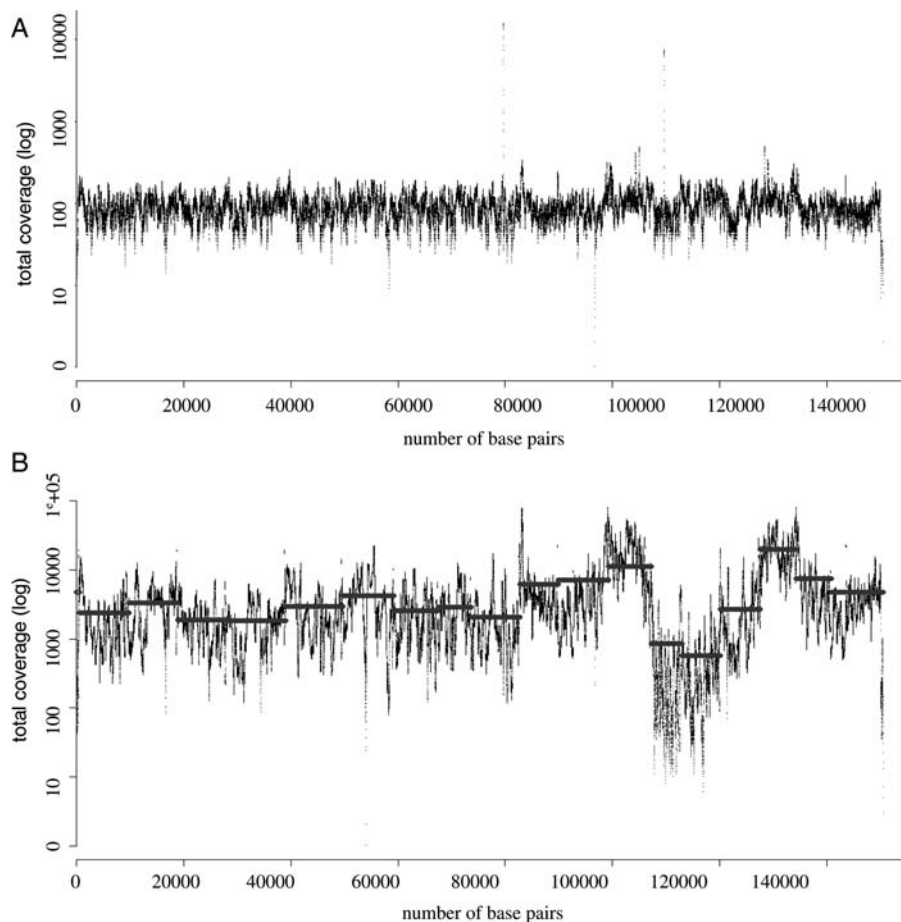

Figure 3. Coverage of the chloroplast genome amplified with a long-range PCR for 16 individuals of *J. vulgaris*.

3.3. Comparison of the chloroplast DNA of J. vulgaris with other Asteraceae genomes analysed

When comparing the full chloroplast genome of *J. vulgaris* with all complete Asteraceae chloroplast genomes (including those from *G. abyssinica*, *H. annuus*, *L. sativa* and *P. argentatum*), a few regions (*trnS-trnC* and *trnE-rpoB*) could not be aligned because these regions were absent in *P. argentatum*, and most other regions showed almost no sequence divergence. Regions that could be aligned and that showed moderate sequence divergence between these five species are listed in Table 2. Five markers (*ndhC-trnV*, *ndhC-atpE*, *rps18-rpl20*, *clpP* and *psbM-trnD*) contained parsimony-informative characters higher than 2% and contained equally high phylogenetic information when compared with other phylogenetic markers that are frequently applied among Asteraceae species such as *trnL-trnF* (6.9%), *trnH-psbA* (1.7%), *rbcL* (1.4%), *rps16* (0.5%) and *ndhF* (0.4%). In Fig. 5, the corresponding single MPTs are depicted.

In a former comparison with *H. annuus* against *L. sativa*[43] and with *H. annuus* against *G. abyssinica*,[44]

Figure 4. (**A**) Whole chloroplast genome coverage plotted for individual 17 of *J. vulgaris*, of which DNA was obtained by using the chloroplast extraction method. (**B**) Whole chloroplast genome coverage plotted for 16 individuals of *J. vulgaris* run in two lanes total, of which DNA was obtained by using the long-range PCR method.

Table 2. Promising regions identified for molecular phylogenetic studies of Asteraceae by comparison of the full chloroplast genomes of *G. abyssinica, H. annuus, Jacobaea vulgaris, L. sativa* and *P. argentatum*

Region	Length (bp)	Tree length	CI	RI	Pars. inf. char. (%)	Topologies gene versus species tree
trnL-trnF[a]	360	100	0.91	0.64	6.9	Incongruent
ndhC-trnV	1189	520	0.89	0.88	4	Congruent
ndhC-atpE	2376	665	0.96	0.75	3.5	Congruent
rps18-rpl20	282	50	0.96	0.78	3	Congruent
clpP	889	181	0.97	0.79	2.6	Incongruent
psbM-trnD	800	114	0.92	0.55	2.5	Incongruent
petN-psbM	569	92	0.97	0.83	2	Congruent
rps8-rps14	219	29	0.96	0.75	2	Incongruent
ycf1	5811	878	0.94	0.59	2	Congruent
ycf3-trnS	1075	232	0.76	0.67	2	Congruent
combined regions	40 449	7719	0.97	0.62	1.8	Congruent
ndhA	2317	208	0.94	0.70	1.7	Congruent
trnH-psbA[a]	1571	172	0.92	0.52	1.7	Congruent
petD	1266	108	0.97	0.86	1.6	Congruent
rbcL[a]	1458	96	0.95	0.76	1.4	Congruent
petB	1490	115	0.96	0.75	1.3	Congruent
ndhI	547	241	0.95	0.83	1	Congruent
rps8-rps3	2451	262	0.94	0.50	1	Congruent
rps15	338	27	0.93	0.50	1	Incongruent
rpoC1	780	82	0.97	0.80	1	Congruent
psbB	1561	78	0.99	0.93	0.8	Congruent
rpoC2	4609	260	0.97	0.81	0.8	Congruent
ndhG	540	31	1.00	1.00	0.7	Congruent
rpoB	3606	133	0.97	0.83	0.6	Congruent
rps16[a]	1159	101	0.99	0.83	0.5	Congruent
cemA	690	47	0.80	0.75	0.4	Congruent
psaC	264	10	1.00	1.00	0.4	Congruent
ndhF[a]	2232	156	0.98	0.67	0.4	Congruent

The CI and RI were calculated with autapomorphic characters excluded.

[a]Commonly used phylogenetic markers included for comparison.

the regions *ndhC-trnV* and *clpP* were already identified as divergent regions within the Asteraceae. CIs of the newly discovered phylogenetic markers, indicating homoplasy, of the newly discovered markers were all in the same range as the commonly used markers except for *ycf3-trnS* and *cemA*, which had slightly lower values. RI values ranged from 0.52 to 0.83 for the commonly used markers and from 0.50 to 1.00 for the newly discovered markers.

Analysis of all 27 regions combined resulted in a congruent topology with high support for all internal nodes. Gene trees can be incongruent with species trees when evolution of genes and species did not occur congruently.[10] Gene trees of five regions (*trnL-trnF, clpP, psbM-trnD, rps8-rps14* and *rps15*) were found to be incongruent with the generally inferred species tree of the Asteraceae species analysed (Table 2; Fig. 5).

With a length of 150 686 bp, *J. vulgaris* has the smallest chloroplast genome compared with the four other Asteraceae cp genomes sequenced so far. The length is 2215 bp less than the largest cp genome of *P. argentatum*. The genome is identical in gene content to *H. annuus* and *L. sativa* and differs in gene number with *G. abyssinica* (which has one gene less) and *P. argentatum* (which has four genes more). Although the similarity in gene content was high, few non-coding regions showed a high sequence divergence between the five Asteraceae species. A number of regions showing sequence divergence between these species contained a high phylogenetic content compared with the standard applied phylogenetic markers used in the Asteraceae. Those regions seem promising for development of universal primers to further investigate clades in molecular phylogenies of Asteraceae hitherto unresolved. Furthermore, many of these regions are not yet used in angiosperm molecular phylogenetic studies[45] and seem worthwhile to investigate further.

3.4. Detection of polymorphic loci

3.4.1. Single-nucleotide polymorphisms The 17 individuals of *J. vulgaris* yielded a total of 32 SNPs (Table 3), which is on average one SNP per 4705 bp. In 66% of the cases, an SNP allele was found only in a single individual. Fifty-nine per cent of the SNP polymorphisms where substitutions from a purine to a pyrimidine or vice versa. No SNPs were found in tRNAs (Table 4). Within the single-copy region (LSC and SSC), SNPs were almost equally divided over coding DNA (tRNA + exons + genic) (13) and intergenic spacers and introns (19). However, in the coding DNA, on average one SNP every 4573 bp was found compared with one SNP on average for every 2780 bp in intergenic and intron spacers (Table 4). Within the genes, two SNPs were located in introns, this is on average one SNP per 3439 bp compared with one SNP per 4811 bp located in coding gene sequences (genes + exons; Table 4). Of the 13 SNPs found in coding DNA, three resulted in non-synonymous substitutions (Table 3).

Reads derived from the IRs are distributed randomly to IRa or IRb by the assembly software. However, if IRa is different from IRb by an indel or SNP, this would be
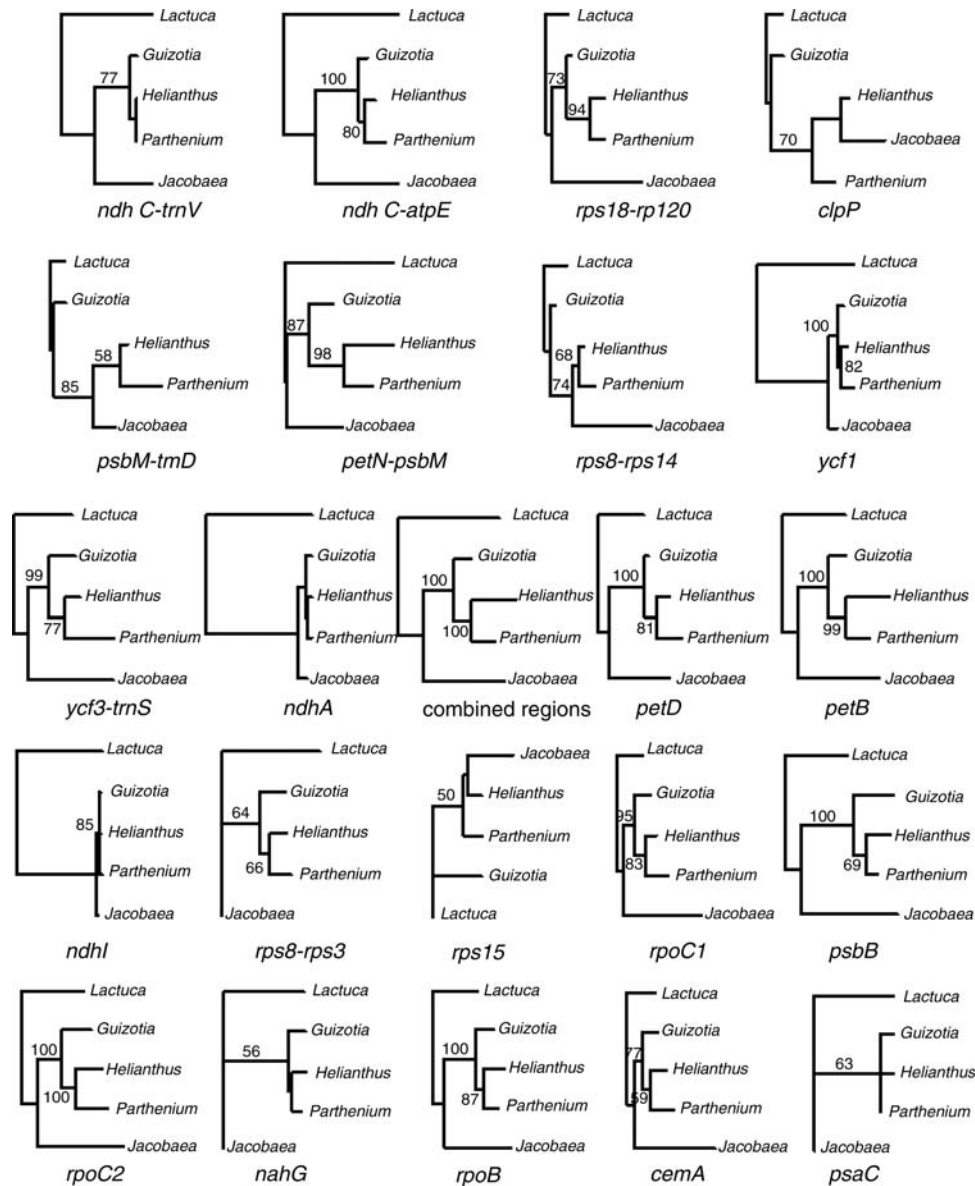
Figure 5. Phylograms derived from maximum parsimony (MP) analysis of alignments of DNA sequences of five different Asteraceae species of a total of 27 individual chloroplast regions indicated below the trees. The phylogram called 'combined regions' in the middle is derived from MP analysis of all 27 regions together.

observed as a polymorphism within an individual. That was, however, never observed as we specifically checked for this. In the one case where we found that the sequence of the IR of individual 11 was deviating from other individuals for four positions, these positions within individual 11 were fully homozygous in both IRa and IRb. All SNPs found in the IRs, 2 × 4 in total, were located in the intergenic spacers of individual 11. The four SNPs found in individual 11 in IRa were found in exactly the same place and the same mutation as in IRb. This suggests 'concerted evolution' or gene conversion for the IR region. On average, one SNP in every 1808 bp was found in the intergenic spacers in the IR. For a subset of 11 SNPs, primers were developed (Table 3) and several

individuals were genotyped using high-resolution melting. For all these individuals, the SNPs were confirmed.

The number of SNPs that were found in this study might be slightly underestimated because the whole cp genome was not mapped with sufficient coverage to detect all SNPs in the 17 individuals analysed. Although the number of synonymous substitutions in chloroplast genes is on average at least three times lower than that of nuclear genes,[24] we still found SNPs using chloroplast genomes of 17 individuals of *J. vulgaris* originating from different populations. We found that SNPs were 1.8 times more frequent in intergenic spacers and introns when compared with DNA coding genes. These findings

Table 3. List of positions and variants of SNPs and microsatellites in the chloroplast genome of *J. Vulgaris*: SNP positions, alleles with the most occurring allele first, frequency of the least occurring allele in 17 individuals of the cp genome of *J. vulgaris* and region and locus of these SNPs

Position	Alleles	Freq.	Region	Locus	Position	Alleles	Freq.	Region	Locus
165	T/A	0.13	Intergenic	*trnH-GUG/psbA*	61 436	C/T	0.31	Genic	*petA*
4032	**A/C**	**0.06**	**Intergenic**	***matK/trnK-UUU***	65 579	G/C	0.06	Intergenic	*trnP-UGG/psaJ*
5555	**A/T**	**0.13**	**Intron**	***rps16***	66 056	T/G	0.19	Intergenic	*psaJ/rpl33*
7837	**A/C**	**0.06**	**Intergenic**	***psbK/psbL***	67 055	G/A	0.25	Intergenic	*Rps18/rpl20*
11 353	**C/A**	**0.06**	**Intergenic**	***trnY-GUA/trnE-UUC***	67 963	T/C	0.69	Intergenic	*Rpl20/rps12*
18 287	**A/C**	**0.13**	**Exon**	***rpoC1***	69 567**	T/C	0.06	Exon	*clpP*
22 648	C/T	0.06	Genic	*rpoC2*	70 234	T/G	0.06	Intron	*clpP*
24 906	T/G	0.38	Intergenic	*atpI-atpH*	92 417*	C/T	0.06	Intergenic	*trnL-CAA/ndhB*
31 299	C/A	0.06	Intergenic	*trnT-GGU/psbD*	97 496*	C/A	0.06	Intergenic	*Rps7/ycf15*
39 790	A/G	0.44	Genic	*psaA*	106 663*	T/G	0.06	Intergenic	*trnR-ACG/trnN-GUU*
39 829	G/A	0.13	Genic	*psaA*	106 664*	C/A	0.06	Intergenic	*trnR-ACG/trnN-GUU*
43 765	C/T	0.06	Intergenic	*Ycf3/trnS-GCA*	108 200**	G/C	0.25	Genic	*Ycf1*
47 181	G/C	0.06	Intergenic	*trnL-UAA/trnF-GAA*	**118 779**	**C/G**	**0.06**	**Genic**	***ndhD***
49 751	C/T	0.06	Genic	*ndhC*	123 423	A/C	0.06	Intergenic	*Rpl32/ndhF*
53 025	G/A	0.06	Genic	*atpB*	124 027	C/T	0.06	Genic	*ndhF*
60 245	**C/T**	**0.06**	**Genic**	***cemA***	124 035**	C/T	0.06	Genic	*ndhF*

SNPs that were tested for multiple individuals with high-resolution melting are indicated by bold typeface.
*SNPs located in the IR.
**Non-synonymous substitutions.

Table 4. Summary of number of basepairs, number of SNPs, number of basepairs per SNP, number of microsatellite regions and number of basepairs per microsatellite region

	nr of bp		nr of SNPs		nr of bp/nr of SNPs		nr of ms		nr of bp/nr of ms	
	SCR	IR	SCR	IR	SCR	IR	SCR	IR	SCR	IR
Non-coding DNA	41 688	8574	16	4	2606	2144	28	1	1489	8574
Coding DNA	59 445	11 688	12	0	4954		5	0	11 889	
rRNA	0	4515	0	0			0	0		
Non-coding gene	6877	1339	2	0	3439		5	0	1375	
Coding gene	57 733	11 174	12	0	4811		6	0	9622	

Non-coding DNA, intergenic spacers and introns; coding DNA, tRNA, genes and exons; non-coding gene, introns; coding gene, genes and exons. All comparisons are made for the single copy region (SCR) and for one IR.

are in line with the assumption that coding DNA generally evolves more slowly than non-coding regions.[30]

The result that individual 11 has four SNPs in both the IR regions suggests that a mechanism is present that provides simultaneous mutations in both IRa and IRb. In all 17 individuals, the sequences of IRa and IRb did not differ from each other by a single base. The gene *Ycf*1 starts at the end of IRb and extends into small single copy (SSC) to yield the full *Ycf*1 sequence. In IRa, the *Ycf*1 gene starts but is not extended into SSC yielding a non-functional sequence. It suggests that there is a selective force that prevents that the IR regions start to deviate from each other even when all the

mutations are located in the intergenic spacer or non-functional genes. As a consequence, the IRs may contribute to the structural stability of the cp genome. Two plant groups, legumes and conifers, lost their IR and comparative sequence studies showed that these chloroplasts experienced a 4-fold increase in silent substitutions compared with chloroplasts containing the IR.[46]

3.4.2. Microsatellites A total of 34 microsatellite regions were found with A/T repeats longer than nine repeats, which is one microsatellite per 4432 bp. Only one microsatellite region was found with 11 G repeats and no repeats of 10 or more Cs were found in the chloroplast genome of *J. vulgaris*.

Table 5. List of positions and variants of SNPs and microsatellites in the chloroplast genome of *J. Vulgaris*: potential microsatellite loci, repeat, repeat length in the consensus chloroplast sequence and the region and locus of these repeats in the cp genome of *J. vulgaris*

Position of repeat	Repeat	Repeat length of consensus	Region	Locus
6705	**A**	**11**	**Intergenic**	**rps16/trnQ-UUG**
12 459	**T**	**14**	**Intergenic**	**trnE-UUC/rpoB**
13 143	A	10	Genic	rpoB
16 413	T	10	Intron	rpoC1
17 759	A	10	Exon	rpoC1
18 185	A	10	Exon	rpoC1
24 848	**A**	**17**	**Intergenic**	**atpL/atpH**
27 760	**T**	**15**	**Intergenic**	**atpF/atpA**
27 776	A	11	Intergenic	atpF/atpA
34 901	A	10	Intergenic	trnS-UGA/psbZ
41 459	**T**	**10**	**Intergenic**	**psaA-ycf3**
41 471	A	13	Intergenic	psaA-ycf3
46 228	A	14	Intergenic	trnT-UGU/trnL-UAA
49 996	G	11	Intergenic	ndhC/trnV-UAC
53 630	A	10	Intergenic	atpB/rbcL
54 013	**T**	**18**	**Intergenic**	**atpB/rbcL**
58 662	T	10	Intergenic	psaL/ycf4
64 247	A	11	Intergenic	psbIE/petL
69 969	A	11	Intron	clpP
70 312	A	10	Intron	clpP
72 916	A	11	Genic/Intergenic	psbT/psbN
74 047	A	11	Intron	petB
76 775	**T**	**17**	**Genic**	**rpoA**
79 191	T	13	Intergenic	rps8/rpl14
79 774	**A**	**10**	**Intergenic**	**rpl14/rpl16**
81 396	T	10	Intergenic	rpl16/rps3
82 909*	T	10	Intergenic	rps19/rpl2
109 743	**A**	**10**	**Genic**	**ycf1**
112 000	A	11	Genic	ycf1
114 539	T	10	Intron	ndhA
121 458	A	11	Intergenic	ccsA/trnL-UAG
121 889	T	11	Intergenic	trnL-UAG/rpl32
123 661	A	10	Intergenic	rpl32/ndhF
150 626*	A	10	Intergenic	rpl2/tnH-GUG

Microsatellites that were tested for polymorphisms in multiple individuals are indicated by bold typeface.
*Microsatellites located in the IR.

Within the single-copy region, 5.6 times as much microsatellite regions were found in intergenic spacers and introns compared with coding DNA (28 against 5, respectively). No microsatellites were found in the tRNA and rRNA. We found on average one microsatellite region every 1489 bp in intergenic spacers and introns against 1 of 11 889 bp in coding DNA (Table 4). Within the genes, microsatellite regions were almost equally divided over exons and genes (6) and introns (5). This is on average 1 of 1375 bp for introns against 1 of 96 222 bp for exons and genes (Table 4). This is not in accordance with the data of SNPs where the number of SNPs per base pair was relatively almost the same for exons + genes and introns. An insertion or a deletion in an exon or gene will lead to a frame shift and therefore likely leads to a non-functional protein. Both microsatellite regions and SNPs occur less in DNA coding regions (exons + genes + tRNA) compared with non-coding regions (intergenic spacers + intron). However, this difference is more marked for microsatellite regions than SNPs.

Of the 34 microsatellite regions, only one was located on IRb in an intergenic spacer. This is surprising because concerted evolution, as earlier suggested, should lead to exact sequence duplication in IRa compared with IRb, and therefore both IRs should contain the same number of nucleotide repeats. Indeed, we found a microsatellite region at the same place on both IRs, but this repeat was only 8 bp on IRa and is therefore not included in Table 5. For 10 repeat regions, primers were developed and multiple individuals from different populations were genotyped (Table 5). Optimization failed for one primer pair, but the other nine regions were amplified and they were all polymorphic. We tested 93 *J. vulgaris* individuals in total and found that all were polymorphic with the number of alleles per locus varying from two to six with an average of 3.3 alleles per locus.

The number of microsatellite regions is promising for investigating allele frequencies in populations and eventually, together with the SNP data, tracing the source population(s) of non-native *J. vulgaris*. The number of variable microsatellites might be higher since we arbitrarily decided to include only mononucleotide repeats that were at least 10 bp long. We found that potential microsatellite regions were 4.7 times more located in intergenic regions and introns compared with coding regions. Because SNPs were only 1.8 times more located in intergenic regions and introns compared with coding regions, we conclude that point mutations are more frequent in coding DNA than indels leading to frame shifts immediately. Although the location of potential microsatellite loci is certain, the repeat length is an approximation. During the long-range PCR and PCR steps in the sample preparation steps for the Illumina platform, indels can occur in microsatellite loci, leading to less or more repeats. Consequently,

the Illumina reads for microsatellite loci differed, making it hard to deduce the repeat length. This could also be the explanation for finding a difference in repeat length of a potential microsatellite locus between the IRs.

In conclusion, we found promising regions for development of universal primers that can be used for further investigation of clades in molecular phylogenies of Asteraceae. Considering the number of SNPs and microsatellites found in this study, we recommend screening of the complete chloroplast genome to find differences within a species. Despite the higher variation in coverage, using the long-range PCR products as templates for Illumina sequencing seemed to be far more efficient than using cpDNA directly.

Supplementary data: Supplementary data are available at www.dnaresearch.oxfordjournals.org.

References

 1. Lundberg, J. and Bremer, K. 2003, A phylogenetic study of the order Asterales using one morphological and three molecular data sets, *Int. J. Plant Sci.*, **164**, 553−78.
 2. Carlquist, S. 1976, Tribal interrelationships and phylogeny of the Asteraceae, *Aliso*, **8**, 465−92.
 3. Hendry, G.A.F. 1993, Evolutionary origins and natural functions of fructans—a climatological, biogeographic and mechanistic appraisal, *New Phytologist.*, **123**, 3−14.
 4. Barker, M.S., Kane, N.C., Matvienko, M., et al. 2008, Multiple paleopolyploidizations during the evolution of the Compositae reveal parallel patterns of duplicate gene retention after millions of years, *Mol. Biol. Evol.*, **25**, 2445−55.
 5. Kim, K.M., Choi, K.S. and Jansen, R.K. 2005, Two chloroplast DNA inversions originated simultaneously during the early evolution of the sunflower family (Asteraceae), *Mol. Biol. Evol.*, **22**, 1783−92.
 6. Julien, M. 1987, *Biological Control of Weeds: A World Catalogue of Agents and their Target Weeds*. CAB International: Wallingford, UK.
 7. Wardle, D.A., Nicholson, K.S. and Rahman, A. 1995, Ecological effects of the invasive weed species *Senecio jacobaea* L (ragwort) in a New Zealand pasture, *Agric. Ecosyst. Environ.*, **56**, 19−28.
 8. Panero, J.L. and Funk, V.A. 2008, The value of sampling anomalous taxa in phylogenetic studies: major clades of the Asteraceae revealed, *Mol. Phylogenet. Evol.*, **47**, 757−82.
 9. Pelser, P.B., Nordenstam, B., Kadereit, J.W. and Watson, L.E. 2007, An ITS phylogeny of tribe Senecioneae (Asteraceae) and a new delimitation of *Senecio* L, *Taxon*, **56**, 1077−104.
10. Pelser, P.B., Kennedy, A.H., Tepe, E.J., et al. 2010, Patterns and causes of incongruence between plastid and nuclear Senecioneae (Asteraceae) phylogenies, *Am. J. Bot.*, **97**, 856−73.
11. Bain, J.F. 1991, Biology of Canadian weeds 96 *Senecio jacobaea* L, *Can. J. Plant Sci.*, **71**, 127−40.
12. Poole, A.L. and Cairns, D. 1940, Botanical aspects of ragwort (*Senecio jacobaea* L.) control, *Bulletin of the New Zealand Dept. of Sci. and Indust. Research*, **82**, 2−61.
13. Schmidl, L. 1972, Biology and control of ragwort, *Senecio jacobaea* L, in Victoria, Australia, *Weed Res.*, **12**, 37−45.
14. Gilkey, H.M. 1957, *Weeds of the Pacific Northwest*, Oregon State College: Corvallis.
15. Witte, L., Ernst, L., Adam, H. and Hartmann, T. 1992, Chemotypes of 2 pyrrolizidine alkaloid-containing *Senecio* species, *Phytochemistry*, **31**, 559−65.
16. Van der meijden, E. and Van der waalskooi, R.E. 1979, Population ecology of *Senecio jacobaea* in a sand dune system 1. Reproductive strategy and the biennial habit, *Journal of Ecology*, **67**, 131−53.
17. McLaren, D. and Mickan, F. 1997, *The Ragwort Management Handbook*. Department of Natural Resources and Environment: Melbourne.
18. Joshi, J. and Vrieling, K. 2005, The enemy release and EICA hypothesis revisited: incorporating the fundamental difference between specialist and generalist herbivores, *Ecol. Lett.*, **8**, 704−14.
19. Keller, S.R. and Taylor, D.R. 2008, History, chance and adaptation during biological invasion: separating stochastic phenotypic evolution from response to selection, *Ecol. Lett.*, **11**, 852−66.
20. Doorduin, L.J., van den Hof, K., Vrieling, K. and Joshi, J. 2010, The lack of genetic bottleneck in invasive Tansy ragwort populations suggests multiple source populations. *Basic Appl. Ecol.*, **11**, 244−50.
21. Estrada-Mena, B., Estrada, F.J., Ulloa-Arvizu, R., et al. 2010, Blood group O alleles innative Americans: implications in the peopling of the Americas, *Am. J. Phys. Anthropol.*, **142**, 85−94.
22. Leniaud, L., Dedeine, F., Pichon, A., Dupont, S. and Bagneres, A.G. 2010, Geographical distribution, genetic diversity and social organization of a new European termite, *Reticulitermes urbis* (Isoptera: Rhinotermitidae), *Biol. Invasions*, **12**, 1389−402.
23. Zimmermann, H., Ritz, C.M., Hirsch, H., Renison, D., Wesche, K. and Hensen, I. 2010, Highly reduced genetic diversity of *Rosa rubiginosa* L. populations in the invasive range, *Int. J. Plant Sci.*, **171**, 435−46.
24. Shendure, J. and Ji, H.L. 2008, Next-generation DNA sequencing, *Nat. Biotechnol.*, **26**, 1135−45.
25. Wolfe, K.H., Li, W.H. and Sharp, P.M. 1987, Rates of nucleotide substitution vary greatly among plant mitochondrial, chloroplast, and nuclear DNAs, *Proc. Natl Acad. Sci. USA*, **84**, 9054−8.

26. Diekmann, K., Hodkinson, T.R., Wolfe, K.H., van den Bekerom, R., Dix, P.J. and Barth, S. 2009, Complete chloroplast genome sequence of a major allogamous forage species, perennial ryegrass (*Lolium perenne* L.), *DNA Res.*, **16**, 165−76.

27. Kumar, S., Hahn, F.M., McMahan, C.M., Cornish, K. and Whalen, M.C. 2009, Comparative analysis of the complete sequence of the plastid genome of *Parthenium argentatum* and identification of DNA barcodes to differentiate *Parthenium* species and lines, *BMC Plant Biol.*, **9**.

28. Wu, F.H., Chan, M.T., Liao, D.C., et al. 2010, Complete chloroplast genome of Oncidium Gower Ramsey and evaluation of molecular markers for identification and breeding in Oncidiinae, *BMC Plant Biol.*, **10**.

29. McCauley, D.E., Stevens, J.E., Peroni, P.A. and Raveill, J.A. 1996, The spatial distribution of chloroplast DNA and allozyme polymorphisms within a population of *Silene alba* (Caryophyllaceae)., *Am. J. Bot.*, **83**, 727−31.

30. Small, R.L., Cronn, R.C. and Wendel, J.F. 2004, Use of nuclear genes for phylogeny reconstruction in plants, *Aust. Syst. Bot.*, **17**, 145−70.

31. Triboush, S.O., Danilenko, N.G. and Davydenko, O.G. 1998, A method for isolation of chloroplast DNA and mitochondrial DNA from sunflower, *Plant Mol. Biol. Rep.*, **16**, 183−9.

32. Doyle, J.J. and Doyle, J.L. 1987, A rapid DNA isolation procedure for small quantities of fresh leaf tissue, *Phytochem. Bull.*, **19**, 11−5.

33. Rozen, S. and Skaletsky, H. 2000, Primer3 on the WWW for general users and for biologist programmers, *Methods Mol. Biol.*, **132**, 365−86.

34. Zerbino, D.R. and Birney, E. 2008, Velvet: algorithms for de novo short read assembly using de Bruijn graphs, *Genome Res.*, **18**, 821−9.

35. Kurtz, S., Phillippy, A., Declcher, A.L., et al. 2004, Versatile and open software for comparing large genomes, *Genome Biol.*, **5**.

36. Wyman, S.K., Jansen, R.K. and Boore, J.L. 2004, Automatic annotation of organellar genomes with DOGMA, *Bioinformatics*, **20**, 3252−5.

37. Conant, G.C. and Wolfe, K.H. 2008, GenomeVx: simple web-based creation of editable circular chromosome maps, *Bioinformatics*, **24**, 861−2.

38. Maddison, W.P. and Maddison, D.R. 2003, *MacClade Version 4.06 OSX*. Sinauer Associates: Sunderland, MA, USA.

39. Swofford, D.L. 2002, *Phylogenetic Analysis Using Parsimony 4.0b*, 10th edition. Sinauer Associated: Sunderland.

40. Felsenstein, J. 1995, *PHYLIP Phylogenetic Inference Programs Version 3.572*. University of Washington.

41. Bao, H., Guo, H., Wang, J.W., Zhou, R.C., Lu, X.M. and Shi, S.H. 2009, MapView: visualization of short reads alignment on a desktop computer, *Bioinformatics*, **25**, 1554−5.

42. Cronn, R., Liston, A., Parks, M., Gernandt, D.S., Shen, R. and Mockler, T. 2008, Multiplex sequencing of plant chloroplast genomes using Solexa sequencing-by-synthesis technology, *Nucleic Acids Res.*, **36**.

43. Timme, R.E., Kuehl, J.V., Boore, J.L. and Jansen, R.K. 2007, A comparative analysis of the *Lactuca* and *Helianthus* (Asteraceae) plastid genomes: identification of divergent regions and categorization of shared repeats, *Am. J. Bot.*, **94**, 302−12.

44. Dempewolf, H., Kane, N.C., Ostevik, K.L., et al. 2010, Establishing genomic tools and resources for *Guizotia abyssinica*; the development of a library of expressed sequence tags, microsatellite loci, and the sequencing of its chloroplast genome, *Mol. Ecol. Resour.*, **10**, 1048−58.

45. Small, R.L., Ryburn, J.A., Cronn, R.C., Seelanan, T. and Wendel, J.F. 1998, Tortoise and the hare: choosing between noncoding plastome and nuclear ADH sequences for phylogeny reconstruction in a recently diverged plant group, *Am. J. Bot.*, **85**, 1301−15.

46. Palmer, J.D. 1990, Contrasting modes and tempos of genome evolution in land plant organelles, *Trends Genet.*, **6**, 115−20.