Understanding the evolution of holoparasitic plants: the complete plastid genome of the holoparasite *Cytinus hypocistis* (Cytinaceae)

Cristina Roquet1,2,* Éric Coissac1,2, Corinne Cruaud3, Marti Boleda1,2, Frédéric Boyer1,2, Adriana Alberti3, Ludovic Gielly1,2, Pierre Taberlet1,2, Wilfried Thuiller1,2, Jérémie Van Es4 and Sébastien Lavergne1,2

1Laboratoire d’Ecologie Alpine, Université Grenoble Alpes, BP 53, FR-38000 Grenoble, France, 2Laboratoire d’Ecologie Alpine, CNRS, BP 53, FR-38000 Grenoble, France, 3CEA-Institut de Génomique, Genoscope, Centre National de Séquençage, FR-91057 Evry Cedex, France and 4Conservatoire Botanique National Alpin, Domaine de Charance, FR-05000 Gap, France

*For correspondence. E-mail cristina.roquet@gmail.com

Received: 2 March 2016 Returned for revision: 9 April 2016 Accepted: 12 May 2016 Published electronically: 21 July 2016

Background and Aims. Plant plastid genomes are highly conserved in size, gene content and structure; however, parasitic plants are a noticeable exception to this evolutionary stability. Although the evolution of parasites could help to better understand plastome evolution in general, complete plastomes of parasites have been sequenced only for some lineages so far. Here we contribute to filling this gap by providing and analysing the complete plastome sequence of *Cytinus hypocistis*, the first parasite sequenced for Malvales and a species suspected to have an extremely small genome.

Methods. We sequenced and assembled de novo the plastid genome of *Cytinus hypocistis* using a shotgun approach on genomic DNA. Phylogenomic analyses based on coding regions were performed on Malvidae. For each coding region present in *Cytinus*, we tested for relaxation or intensification of selective pressures in the *Cytinus* lineage compared with autotrophic Malvales.

Key Results. *Cytinus hypocistis* has an extremely divergent genome that is among the smallest sequenced to date (19-4 kb), with only 23 genes and no inverted repeat regions. Phylogenomic analysis confirmed the position of *Cytinus* within Malvales. All coding regions of *Cytinus* plastome presented very high substitution rates compared with non-parasitic Malvales.

Conclusions. Some regions were inferred to be under relaxed negative selection in *Cytinus*, suggesting that further plastome reduction is occurring due to relaxed purifying selection associated with the loss of photosynthetic activity. On the other hand, increased selection intensity and strong positive selection were detected for rpl22 in the *Cytinus* lineage, which might indicate an evolutionary role in the host–parasite arms race, a point that needs further research.

Key words: Chloroplast genome, Cytinaceae, Cytinus hypocistis, Malvales, parasite, mycoheterotroph, plastome evolution, selective pressure.

INTRODUCTION

The chloroplast genome (i.e. plastome) of seed plants is well known for being remarkably conserved in size, structure and gene content across the tree of life (Jansen and Ruhlman, 2012). The most striking exceptions to the evolutionary stability of the plastome have been described in parasite species (Westwood et al., 2010), which show a clear tendency towards plastome reduction (~11–100 kb) in comparison with their autotrophic relatives (usually 120–170 kb; Ruhlman and Jansen, 2014). This reduction is more pronounced in holoparasites, which are incapable of photosynthesis and thus completely dependent on their hosts for their nutrition. For instance, the plastome of the endoparasite *Pilostyles aethiopica* (Apodanthaceae) is the smallest reported to date for land plants, with 11 kb and five possibly functional genes (Bellot and Renner, 2016). In Orobancheaeae, the most studied plant parasite lineage to date, some mechanisms of DNA retention have been suggested (Wicke et al., 2013), but our understanding of plastome evolution in parasites remains limited because complete plastome sequences have not been reported for some parasite lineages, and for others only one or few species have been studied. This gap is unfortunate, as it would permit a more general understanding of the evolutionary forces that shape plastomes.

Evolutionary forces driving plastome reduction

Since the main function of chloroplasts is photosynthesis, a major question is why a plastid genome is maintained in non-photosynthetic plants. Barbrook and colleagues (2006) proposed the ‘essential tRNAs’ hypothesis, which is based on the observation that the plastid trnE is involved in the synthesis of tetrapyrroles (i.e. molecules required for vital processes such as cellular respiration) and cannot be functionally replaced by any cytosolic tRNA for this function (Kumar et al., 1996). Therefore, even non-photosynthetic plants would need to retain trnE (along with the machinery for its transcription), in order to be able to synthesize essential tetrapyrroles such as the haem component of mitochondrial cytochromes and other essential
oxidative enzymes (Barbrook et al., 2006). However, two recent studies based on high-throughput DNA sequencing affirmed that plastid genomes had vanished at least in two non-photosynthetic organisms – the parasitic plant Rafflesia lagascae (Molina et al., 2014) and the colourless alga Polytomella (Smith and Lee, 2014) – and a third study stated that no functional tRNAs were found in Pilostyles (Bellot and Renner, 2016). These plastome losses would mean that the essential tRNAs hypothesis does not apply for these species, and that an alternative pathway exists for haem biosynthesis (Smith and Asmail, 2014).

Another question related to plastome maintenance in parasites is why different subsets of genes are retained in the plastome of different parasitic species (Fig. 1), even within the same genus (Supplementary Data Table S1). Within Cuscuta (Convolvulaceae), plastome differences of the four species studied to date seem to be due to differences in their photosynthetic activity (Funk et al., 2007; McNeal et al., 2007). For Orobanchaceae, the whole plastome has been reported for 16 holoparasites and one hemiparasite (Wolfe et al., 1992; Li et al., 2013; Wicke et al., 2013; Cusimano and Wicke, 2015), most of them showing notable differences in plastid gene content and length and structural rearrangements. A recent study (Wicke et al., 2013) indicated that the retention of plastid DNA fragments in this family could be affected by both the proximity to genes under selection and their co-occurrence with those genes in operons. Regarding Santalales (a large order of plants consisting almost entirely of hemiparasites), only four species have had their plastome sequenced; the most prominent plastome modifications were found in the species with the highest nutritional dependence on the host (Petersen et al., 2015). Concerning mycoheterotrophs (i.e. species that parasitize other plants using fungi as vectors), several monocot lineages have been studied, showing very different degrees of plastome reduction, especially in orchids (Delannoy et al., 2011; Logacheva et al., 2011, 2014; Barrett and Davis, 2012; Braukmann and Stefanovic, 2012; Barrett et al., 2014; Lam et al., 2015; Schelkunov et al., 2015). In summary, all these studies converge in showing that the plastome structure of parasitic plants is often radically altered in comparison with photosynthetic plants, with extensive gene losses and rearrangements in gene order. However, parasitism has independently appeared at least 12 times in seed plants (Barkman et al., 2007; Su et al., 2015), and the fact that the entire plastome is still not known for several parasite lineages limits our understanding of the macroevolutionary context of the emergence of parasitic lifestyles.

A new insightful case study in Cytinaceae

Our study focuses on Cytinus hypocistis, which belongs to Cytinaceae. Cytinus was long considered to belong to
Rafflesiaceae; Bouman and Meijer (1994) highlighted morphological differences in the reproductive traits of Cyttinus compared with other members of this family and described a new tribe (Cytinaceae), but maintained the familial placement. Later, molecular analyses supported the placement of Cyttinus in a separate family (Cytinaceae) within Malvales (Nickrent et al., 2004; Barkman et al., 2007; Nickrent, 2007). However, these phylogenetic studies were only based on two mitochondrial regions (matR, ap1) and the nuclear 18S rDNA sequence for Cyttinus. This plant family with only two genera (Cyttinus and Bdallophytum) shows one of the most extreme manifestations of parasitism: holoparasitic species whose vegetative body is reduced to an endophytic system living within its host root, which only emerges during the blooming period when flowers arise from host tissues. The genus Cyttinus contains about six species, distributed in the Mediterranean Basin, South Africa and Madagascar. In the Mediterranean Basin, two species are commonly distinguished, C. hypocistis and C. ruher; however, a recent study showed that C. ruber is no more differentiated genetically than three subspecies of C. hypocistis (de Vega et al., 2008). Cyttinus hypocistis parasitizes exclusively roots of Cistaceae (Maire, 1961; de Vega et al., 2010) with a striking specialization at the host level, with distinct genetic races associated with different host plant species (De Vega et al., 2008). This species is monoeccious and presents short spikes of basal female flowers and distal male flowers that are mostly anpollinated (De Vega et al., 2009).

Here, we provide the first complete plastid genome sequence for a member of Cytinaceae (Cyttinus), built from Illumina short-read sequencing data using a shotgun approach on genomic DNA. Our first objective was to confirm the phylogenetic placement of Cytinaceae. Second, we aimed to identify which genes have been lost or show evidence of pseudogene formation compared with their autotrophic relatives. More specifically, we looked for the inverted repeat regions using RepSeek (Achaz et al., 2013). The assembled sequence has been deposited in the Sequence Read Archive (accession number PRJEB9903).

De novo assembly and annotation of the complete plastome of C. hypocistis

The sequence was assembled de novo into a single circular contig using ORG.Asm, an assembler specially developed for the assembly of organelle genomes (available at http://metabar coding.org/asm-asn), based on a De Bruijn graph, and already used for the assembling of several plastomes (e.g. Besnard et al., 2014; Malé et al., 2014; Colli et al., 2015; Kocher et al., 2015). To assess the quality of the reconstructed sequence, raw reads were mapped against the reconstructed sequence with the software BWA (Li and Durbin, 2009). The mean coverage of the reconstructed sequence was 138 x, ranging from 78 x to 350 x. Pair-end information was consistent with the assembled sequence, and the insert size was in accordance with library specifications.

Protein-encoding genes were identified based on the detection of open reading frames. For genes that included an intron or, like rps12, were trans-spliced, boundaries of exons were defined based on similarities with already annotated genes available in GenBank. Functional annotations of the encoded proteins were assigned by similarities and manually curated. The 23S, 16S and 5S rRNA genes were identified using rRNAnamer (Lagesen et al., 2007), the 4.5S rRNA gene was identified using BLAST by searching against a local RNA database constituted by a set of 4.5 rRNA genes extracted from complete plastomes downloaded from the NCBI ftp site (http://ftp.ncbi. nlm.nih.gov/refseq/release/plastid/). Transfer RNA genes were annotated with tRNAscan-SE (Lowe and Eddy, 1997). We looked for the inverted repeat regions using RepSeek (Achaz et al., 2007). The plastid genome map was drawn with OGDWdraw (Lohse et al., 2013). The assembled sequence has been deposited in GenBank under the accession number KT335971.

Phylogenomic analyses

In order to confirm the phylogenetic position of Cyttinus within Malvales (as proposed by Nickrent et al., 2004) and obtain a phylogenetic tree for molecular evolution analyses (see below), we recovered all chloroplast genome sequences available in GenBank for all Malvidae s.l. (i.e. including Geraniiales, Myrtales, Crossosomatales, Sapindales, Brassicales and Malvales) plus several levels of outgroup taxa: three rosid species belonging to Malpighiales, Fabales and

MATERIALS AND METHODS

DNA extraction and sequencing

One sample of Cyttinus hypocistis was collected in south-eastern France (44.326894° N, 4.819041° E) by Jérémie Van Es (voucher deposited at the herbarium of the Station Alpine Joseph Fourier [GR]). Total genomic DNA was extracted from silica-dried material of one individual using the DNeasy Plant Kit (Qiagen) and subsequently quantified using the NanoDrop spectrophotometer. Genomic DNA (250 ng) was used for illumina library preparation based on a semi-automated protocol using the SPIRIT Works Library Preparation System and a SPIR TE instrument (Beckmann Coulter), according to the manufacturer’s instructions. DNA fragments were amplified by 12 PCR cycles using a Platinum Pfx Taq Polymerase Kit (Life Technologies) and Illumina adapter-specific primers, and size-selected on 2 % agarose gel (~600 bp). After library profile analysis conducted using an Agilent 2100 Bioanalyzer (Agilent Technologies, USA) and qPCR quantification (MxPro, Agilent Technologies, USA), the library was sequenced using 101 base-length read chemistry in a paired-end flow cell on the Illumina HiSeq2000 sequencer (Illumina, USA). The raw sequences have been deposited in the Sequence Read Archive (accession number PRJEB9903).
Roquet et al. — The plastid genome of the holoparasite Cyttinus hypocistis (Cytinaceae)

Rosales (Populus alba, Lotus japonicus and Pentactina rupicola, respectively), one species of Saxifragales (Sedum sarmentosum) and one species from Vitales (Vitis rotundifolia). Sequences were downloaded on 17 June 2014. When more than one genomic sequence was available for a species, we kept only one (species list and accession numbers are provided in Supplementary Data Table S2). We then extracted coding regions with OBITools (Boyer et al., 2015). We retrieved all coding regions and kept only those that were present at least in 50 % of all study species, resulting in 78 regions. When a coding region was not present, it was coded as missing data.

Codon-based alignments of each coding region were generated with MACSE (Ranwez et al., 2011). MACSE is able to handle frameshifts in the reading frames, which can be artefacts (e.g. sequencing errors) or of biological origin (it has been shown that reading frame shifts can be tolerated during translation; Farabaugh, 1996). Resulting alignments were visually checked. Poorly aligned regions were removed using Gblocks (Castresana, 2000) with the ‘codon’ option activated (i.e. selected blocks contain only complete codons). To check for anomalous sequences, we built maximum likelihood (ML) trees from each of the alignments with RAxML v.8 (Stamatakis, 2014), partitioning first, second and third codon positions separately, and performed bootstrap analyses with the automatic stopping criteria implemented with the option autoMRE. We detected two sequences in obvious spurious phylogenetic locations (rps2 in Lobularia maritima and rps12 in Arabis alpina). We removed those anomalous sequences from the alignments and realigned them following the steps described above. Definitive alignments were concatenated using FASconCAT (Kück and Meusemann, 2010). We used the ‘greedy’ algorithm implemented in PartitionFinder v.1.1.1 (Lanfear et al., 2012) to choose the optimal partition scheme (according to the Bayesian information criterion) for phylogenetic inference starting from an initial partition scheme based on gene and codon position. We ran 200 ML searches from different starting trees using RAxML with the concatenated dataset, applying the GTRCAT model, the rapid hill-climbing algorithm (Stamatakis et al., 2008) and the partition scheme selected with PartitionFinder (21 partitions). Bootstrap searches (1000 replicates) were executed separately using the standard bootstrap option in RAxML. Bootstrap results were drawn in the best ML tree obtained in the previous searches (Fig. 2). The same analyses were performed with a DNA matrix including only the coding regions present in Cyttinus, with the aim of checking whether the missing data for Cyttinus (i.e. genes that have been lost) have an impact on phylogenetic inference.

Test for selective pressures and estimation of substitution rates

For each potential coding region, we tested for potential relaxed or intensified selection in Cyttinus compared with related autotrophic taxa using the hypothesis-testing framework RELAX (Wertheim et al., 2015) available on the Datamonkey webserver (www.datamonkey.org/RELAX). Importantly, RELAX overcomes the statistical limitations of existing methods aiming to identify targets of natural selection. Most of these methods are based on the estimation of the ratio of non-synonymous (dN) and synonymous (dS) substitutions (\(\omega = \text{dN/dS}\)) to test for deviations of the neutral expectation \(\omega = 1\). RELAX tests whether selection is relaxed or intensified in a subset of predefined test branches in a tree (compared with reference branches) with the selection intensity parameter \(k\), which distinguishes whether selective strength on the test subset of branches is compressed towards or repelled away from neutrality compared with the reference subset. In the RELAX null model, \(k\) is constrained to 1 for all branches, whereas in the alternative one it is allowed to differ between reference and test branches. A likelihood-ratio test is implemented to accept or reject the alternative model, and values of the Akaike information criterion with correction for finite sample size (AICc) are also provided. For both models, three categories of \(\omega\) are estimated (one for sites under purifying selection, \(\omega \ll 1\); another for nearly neutral ones, \(\omega \approx 1\); and a third for sites under diversifying selection, \(\omega \gg 1\)), together with the relative proportion of sites they represent. Additionally, RELAX fits the partitioned MG94xREV model for descriptive purposes and to provide a goodness-of-fit reference for the hypothesis-testing framework: in this model separate \(\omega\) distributions are estimated for test and reference branches without accounting for the selection intensity parameter.

RELAX requires \textit{a priori} specification of the test and reference branches. Because our \textit{a priori} hypothesis was that some coding regions might have experienced a different selective pressure in Cyttinus because of parasitism, we set the Cyttinus branch as the test one. All RELAX analyses were run only with Malvales (plus an outgroup from Brassicales, Carica papaya) because exploratory analyses with the PAML (phylogenetic analysis by maximum likelihood) package (Yang, 2007; results not shown) with all Malvidae taxa included in the phylogenetic analyses yielded very high values of \(\omega\) in very short branches, which is an indication that too divergent sequences are being compared. The analyses were run providing as input the best ML tree obtained with RAxML, as described in the previous section, after pruning all non-Malvales taxa except Carica.

As a complement to RELAX analyses, we also fitted site and branch-site codon models implemented in the software CODEML included in the package PAML v.4.8 (Yang, 2007). These models estimate rates of synonymous and non-synonymous substitutions (dS and dN) and the \(\omega\) ratio (dN/dS), taking into account potential rate differences in individual codon sites (site models) and also in specific lineages (branch-site models). Full methodological details are provided in Supplementary Data Methods.

RESULTS

Gene content and organization of C. hypocistis plastome

The complete plastid genome of C. hypocistis is 19 400 bp in length, and it encodes a total of 26 genes, including 13 protein-coding genes (5501 bp), six tRNAs (453 bp) and four rRNAs (4573 bp) (Table 1, Fig. 3). Nucleotides corresponding to coding regions account for 28.4 % of the genome. Three additional genes (cblP, rps2 and rps7) show strong sequence divergence compared with other Malvales and could therefore be pseudogenes; we thus did not include them in the set of coding regions for subsequent analyses. Three other regions have certainly become pseudogenes (and hence were also excluded from
subsequent analyses): \textit{rpl32} encodes only the first 28 amino acids; \textit{accD} lacks the sequence encoding the 200 first amino acids; and in \textit{rps18} the first 20 amino acids are replaced by 30 others and the last 20 amino acids are missing.

The reconstructed genome lacks inverted repeat (IR) copies. The overall GC content of the whole genome is 29.19\%. This value is slightly lower in protein-encoding genes (29.13\%) and much lower in intergenic spacer regions (16.9\%), while tRNA and rRNA show higher GC values, with 51.4\% and 47.5\% respectively. The six tRNAs found in the plastid genome match six different codons corresponding to six amino acids: cysteine, glutamic acid, glutamine, isoleucine, N-formylmethionine and tyrosine.

**Phylogenomic analysis**

The final concatenated alignment used for analyses was 58 863 bp in length and corresponded to 78 coding regions (of which only 13 were present as likely coding genes for \textit{Cytinus}). The proportion of gaps and undetermined characters was 6-24\% (5.39\% excluding \textit{Cytinus}). The likelihoods of the 200 ML trees obtained varied only slightly (424 574 - 424 575 for the best ML tree). The bootstrap analysis showed that most of the nodes were highly supported (86–100\% bootstrap support [BS]), five nodes received strong support (71–85\% BS), five received moderate support (50–70\% BS) and only four did not receive any support (<50\% BS). The nodes with moderate or no support corresponded to intrageneric relationships for \textit{Gossypium} and \textit{Eucalyptus} taxa, except for the node relating the genus \textit{Lepidium} to \textit{Nasturtium} and \textit{Barbarea} taxa (Brassicaceae) (99\% BS). Phylogenetic analyses provided strong support (84\% BS) for a sister relationship between \textit{Thymelaeaceae} and \textit{Malvaceae}, and for a \textit{Malvales} clade also containing \textit{Cytinaceae} (99\% BS, Fig. 2). The \textit{Malvales} clade was sister to the clade corresponding to \textit{Brassicales} (100\%)}
BS). Phylogenetic analyses based on the reduced matrix (only including the coding genes present in Cytinus) yielded congruent results and similar branch lengths compared with the complete DNA matrix (Supplementary Data Fig. S1).

Signatures of selection in plastid genes and substitution rates

For each putative coding region found in the Cytinus plastome, we tested whether different selective pressures can be detected on Cytinus compared with autotrophic Malvales. We did so by comparing two models implemented in RELAX, which makes it possible to investigate whether the selection intensity (k) is different between the test (here, Cytinus) and reference branches. Model parameter estimates are summarized in Table 2. For the majority of the regions, we did not find significant differences in selection patterns between Cytinus and other Malvales, i.e. the null model (which constrains the selection intensity to 1 for all branches) could not be rejected, and a low ω ratio was estimated for the majority of the sites (Table 2). The null model was rejected in favour of the alternative one only for five regions: rpl16, rps3, rps11, rps19 and rpl22. The region rpl22 showed a significant high increase in selection intensity (k ≫ 1, P = 0.04) for the Cytinus lineage, with a small proportion of its sites (6 %) under positive selection in Cytinus while evolving neutrally in other Malvales. In contrast, the other four regions showed a relaxation of selection (k < 1).

Specifically, we found weaker purifying selection for a notable proportion of sites of these genes, with ω values that are higher (and nearer to the neutral expectation of ω = 1) in Cytinus than in Malvales.

The fitted site and branch-site models implemented in PAML provided congruent results with those obtained with RELAX concerning selective constraints on each putative coding plastid region found in Cytinus. A summary of the compared models and the parameter estimates of the best model according to a likelihood ratio test are provided in Supplementary Data Table S3. For rpl22, the branch-site model MA fitted better than the null MAn (P = 0.0172), which indicates that some sites (specifically, 9 %) evolve under positive selection only on the Cytinus branch. For four regions (rps11, rps3, rps19 and rpl16) a branch-site model also fitted better than site models according to the AIC (Supplementary Data Table S4), with a considerable proportion of sites inferred to be under neutral evolution in the Cytinus branch (17–39 %) while evolving under strong negative selection in the other lineages. For all other regions, the null models provided a better fit to the data, and the low ω ratio estimated for the majority of the sites indicated that these genes are evolving under purifying selection. Synonymous (dS) and non-synonymous (dN) substitution rate values estimated for Cytinus and all Malvales are provided in Supplementary Data Table S5.

### DISCUSSION

#### General features of the C. hypocistis plastome

The Cytinus plastome reported here represents the first complete sequence for one of the 12 clades where parasitism evolved independently in angiosperms. This genome shows a spectacular reduction to only 19.4 kb, confirming previous indications from Southern blots of digested genomic DNA, in which plastid gene probes hybridized to a fragment of ~20 kb in length (Nickrent et al., 1997b). This plastome is extremely reduced both in size and gene content, with only 16 protein-coding genes (all corresponding to 50S or 30S ribosomal proteins, except clpP, which encodes ATP-dependent protease), three pseudogenes, four rRNAs and six tRNAs (Table 1). Three of the potentially coding genes are highly divergent and thus need further confirmation from empirical research. To date, the smallest known plastomes correspond to the endoparasitic Apodanthaceae Pilostyles aethiopica and Pilostyles hamiltonii (five genes and 11–15 kb, respectively; Bellot and Renner, 2016), followed by the mycoheterotrophic orchids Epipogium roseum (19 kb and 29 genes) and Epipogium aphyllum (31 kb and 27 genes; Schelkunov et al., 2015), the strange holoparasite Hydnora visseri (27 kb and 24 genes; Naumann et al., 2016), the parasitic green alga Helicosporidium sp. (37.5 kb and 54 genes; De Koning and Keeling, 2006) and the Orobanchaceae Conopholis americana (45.5 kb and 43 genes; Wicke et al., 2013).

Compared with the closest photosynthetic relatives of Cytinus, ~80 % of plastid genes have been lost, including all genes required for photosynthesis, those coding for the plastid RNA polymerase (PEP) and the maturase-like protein matK, seven genes encoding ribosomal proteins and all but six tRNAs. However, we cannot exclude the possibility that some missing
genes have been transferred to the nuclear genome, as the low coverage of this genome in our data makes it impossible to test this scenario. It is particularly intriguing why only six tRNAs were retained in *Cytinus* while the rest were lost. The tRNAs present in *Cytinus* are found in all parasites sequenced to date, with the following exceptions: *trnC*-GCA is missing in *Conopholis*, *Hydnora* and *Phelipanche purpurea*; *trnQ*-UUG is not present in *Epipogium aphyllum* and *Hydnora*; and *trnY*-GUA is not found in *Sciaphila* and *Hydnora* (Supplementary Data Table S1). In other words, half of the tRNAs conserved in *Cytinus* (*trnE*-UUC, *trnfM*-CAU, *trnl*-CAU) are present in almost all sequenced parasitic and fully mycoheterotrophic plants (no functional tRNAs have been found in *Pilostyles*), and it is noteworthy that they correspond to the three unique tRNAs found in *Hydnora* (Naumann et al., 2016). Among them, two of the three tRNAs carrying the CAU anticodon are kept. They load the isoleucine and formyl-methionine needed to ensure correct translation initiation without conflict with the isoleucine codon translation in chloroplasts and bacteria (Kashdan and Dudock, 1982). The third tRNA (CAU) dedicated to the classical methionine amino acid is absent. It has been proposed that nuclear-encoded tRNAs might be imported from the cytosol into chloroplast to compensate for missing genes (Morden et al., 1991), and indeed this has been shown experimentally to occur in plant mitochondria (e.g. Dietrich et al., 1992), but not in plastids so far. If this was the case, the question then arises as to whether the remaining tRNAs are maintained in the plastome because they have not yet been subjected to deletion, or whether a few specific tRNAs have an essential function and cannot be imported from the cytoplasm for some particular reasons, as has been proposed in the essential tRNAs hypothesis (Barbrook et al., 2006). Our results (together with data on the other known parasitic plastomes to date) point to the latter possibility, despite the notable exceptions of *Rafflesia lagascae* and the alga *Polytomella*, in which the plastome is apparently entirely absent (Molina et al., 2014; Smith and Lee, 2014), and
that the reduction of the GC content in the plastome coincides species usually have 35–40% GC content). It has been shown parasites belonging to different lineages (Nickrent 2013; Martinez-Alberola 2011). Inverted repeats are thought to help stabilize the plastid that could result from plastome rearrangements related to the dele-

tions correspond to truncated coding regions. These areas could result from plastome rearrangements related to the dele-

Table 2. Test for selective pressures on plastid genes of Cytinus compared with its Malvales autotrophic relatives (‘reference group’). The parameter k is the estimated selection intensity; values under and above 1 indicate relaxed and intensified selection in Cytinus, respectively. The P value is indicated in bold for cases in which the null model was rejected (P < 0.05). Columns $\omega_{\text{Cyti}}$ and $\omega_{\text{ref}}$ indicate $\omega$ values calculated for Cytinus and reference branches under the MG94xREV model, respectively. Columns $\omega_1$, $\omega_2$ and $\omega_3$ indicate three categories of $\omega$ values ($\omega_1$ for sites under purifying selection; $\omega_2$ for nearly neutral ones; $\omega_3$ for sites under diversifying selection) estimated with the alternative or the null model depending on which is most suitable according to P value; values in parentheses correspond to the relative proportion of sites they represent.

<table>
<thead>
<tr>
<th>Gene</th>
<th>k</th>
<th>P</th>
<th>AICcNull</th>
<th>AICcAlt.</th>
<th>$\omega_1^{\text{Cyti}}$</th>
<th>$\omega_1^{\text{ref}}$</th>
<th>$\omega_1^{\text{Cyti}}$</th>
<th>$\omega_1^{\text{ref}}$</th>
<th>$\omega_2^{\text{Cyti}}$</th>
<th>$\omega_2^{\text{ref}}$</th>
<th>$\omega_3^{\text{Cyti}}$</th>
<th>$\omega_3^{\text{ref}}$</th>
</tr>
</thead>
<tbody>
<tr>
<td>rpl2</td>
<td>0.96</td>
<td>0.90</td>
<td>3234.1</td>
<td>3236.2</td>
<td>0.25</td>
<td>0.27</td>
<td>0.17 (84%)</td>
<td>0.17 (84%)</td>
<td>0.7 (16%)</td>
<td>0.7 (16%)</td>
<td>–</td>
<td>–</td>
</tr>
<tr>
<td>rpl14</td>
<td>0.85</td>
<td>0.44</td>
<td>1718.1</td>
<td>1719.6</td>
<td>0.15</td>
<td>0.09</td>
<td>–</td>
<td>–</td>
<td>0.10 (100%)</td>
<td>0.10 (100%)</td>
<td>–</td>
<td>–</td>
</tr>
<tr>
<td>rpl16</td>
<td>0.47</td>
<td>0.04</td>
<td>1925.2</td>
<td>1923.4</td>
<td>0.20</td>
<td>0.08</td>
<td>0.57 (57%)</td>
<td>0.57 (57%)</td>
<td>0.44 (43%)</td>
<td>0.18 (43%)</td>
<td>–</td>
<td>–</td>
</tr>
<tr>
<td>rps20</td>
<td>0.06</td>
<td>0.15</td>
<td>1953.9</td>
<td>1954.2</td>
<td>0.22</td>
<td>0.24</td>
<td>&lt;0.01 (64%)</td>
<td>&lt;0.01 (64%)</td>
<td>0.73 (36%)</td>
<td>0.73 (36%)</td>
<td>–</td>
<td>–</td>
</tr>
<tr>
<td>rpl22</td>
<td>49.5</td>
<td>0.04</td>
<td>1089.3</td>
<td>1087.6</td>
<td>0.47</td>
<td>0.33</td>
<td>0.50 (50%)</td>
<td>0.50 (50%)</td>
<td>1.44 (44%)</td>
<td>1.44 (44%)</td>
<td>337 (6%)</td>
<td>1.12 (6%)</td>
</tr>
<tr>
<td>rpl36</td>
<td>1.06</td>
<td>0.94</td>
<td>571.4</td>
<td>574.1</td>
<td>0.12</td>
<td>0.12</td>
<td>&lt;0.01 (49%)</td>
<td>&lt;0.01 (49%)</td>
<td>0.21 (51%)</td>
<td>0.21 (51%)</td>
<td>–</td>
<td>–</td>
</tr>
<tr>
<td>rps3</td>
<td>0.48</td>
<td>0.02</td>
<td>3555.8</td>
<td>3552.4</td>
<td>0.27</td>
<td>0.14</td>
<td>&lt;0.01 (39%)</td>
<td>&lt;0.01 (39%)</td>
<td>0.47 (61%)</td>
<td>0.21 (61%)</td>
<td>–</td>
<td>–</td>
</tr>
<tr>
<td>rps4</td>
<td>23.4</td>
<td>0.08</td>
<td>1599.7</td>
<td>1598.9</td>
<td>0.29</td>
<td>0.20</td>
<td>0.76 (76%)</td>
<td>0.76 (76%)</td>
<td>–</td>
<td>–</td>
<td>1.4 (24%)</td>
<td>1.4 (24%)</td>
</tr>
<tr>
<td>rps8</td>
<td>0.2</td>
<td>0.08</td>
<td>2305.0</td>
<td>2304.9</td>
<td>0.29</td>
<td>0.3</td>
<td>0.34 (34%)</td>
<td>0.34 (34%)</td>
<td>0.33 (64%)</td>
<td>0.33 (64%)</td>
<td>9.37 (2%)</td>
<td>9.37 (2%)</td>
</tr>
<tr>
<td>rps11</td>
<td>0.2</td>
<td>&lt;0.01</td>
<td>1968.6</td>
<td>1958.5</td>
<td>0.29</td>
<td>0.05</td>
<td>0.08 (67%)</td>
<td>&lt;0.01 (67%)</td>
<td>0.68 (33%)</td>
<td>0.14 (33%)</td>
<td>–</td>
<td>–</td>
</tr>
<tr>
<td>rps12</td>
<td>1.36</td>
<td>0.49</td>
<td>1378.2</td>
<td>1379.9</td>
<td>0.18</td>
<td>0.29</td>
<td>–</td>
<td>–</td>
<td>0.17 (100%)</td>
<td>0.17 (100%)</td>
<td>–</td>
<td>–</td>
</tr>
<tr>
<td>rps14</td>
<td>0.65</td>
<td>0.12</td>
<td>1521.1</td>
<td>1520.9</td>
<td>0.24</td>
<td>0.11</td>
<td>–</td>
<td>–</td>
<td>0.17 (100%)</td>
<td>0.17 (100%)</td>
<td>–</td>
<td>–</td>
</tr>
<tr>
<td>rps19</td>
<td>0.25</td>
<td>0.02</td>
<td>1298.4</td>
<td>1294.9</td>
<td>0.19</td>
<td>0.06</td>
<td>0.01 (73%)</td>
<td>&lt;0.01 (73%)</td>
<td>0.65 (27%)</td>
<td>0.19 (27%)</td>
<td>–</td>
<td>–</td>
</tr>
</tbody>
</table>

two endoparasitic species of Pilostyles (Apodanthaceae) for which no trace of functional tRNAs has been found in their extremely reduced plastome (Bellot and Renner, 2016). Regarding its genomic structure, the plastome of C. hypocistis does not present the typical IR regions, resulting in the inability to define large single copy (LSC) and small single copy (SSC) regions. Absence of IRs has been reported for the parasites Conopholis americana and Phelipanche ramosa (Orobanchaceae), and also for certain non-parasites, including a big monophyletic clade of Fabaceae (the ‘inverted repeat-lacking’ clade; Wojciechowski et al., 2000), certain conifers (Strauss et al., 1988) and a few genera of Ericaceae (Fajardo et al., 2013; Martinez-Alberola et al., 2013), Poaceae (Guisinger et al., 2010) and Geraniaceae (Guisinger et al., 2011). Inverted repeats are thought to help stabilize the plastid genome structure (Palmer and Thompson, 1982), and their disappearance might accelerate gene loss for regions that are not functionally important any more due to parasitism. Further work on the comparative structure of plastomes in relation to species divergence time from autotrophic relatives should allow the identification of the parts of the plastome that are lost first following the loss of autotrophy, and in particular whether the loss of IRs triggers plastome reduction. The Cytinus plastome also presents repeated DNA motifs. In particular, we found a motif of ~100 bp repeated six times with a similarity of 70–90%, and some of these repetitions are contiguous to sequences that could correspond to truncated coding regions. These areas could result from plastome rearrangements related to the deletion of some parts of the genome. The C. hypocistis plastome possesses a relatively low GC content of 29.9% (non-parasitic species usually have 35–40% GC content). It has been shown that the reduction of the GC content in the plastome coincides with the transition to heteroorth and increases gradually in holoparasitic Orobanchaceae (Wicke et al., 2013); low GC content was also found in plastid-derived 16S rRNA for seven holoparasites belonging to different lineages (Nickrent et al., 1997a). These results suggest a relationship between GC downward bias and relaxation of selective pressures linked to the loss of photosynthesis.

Selective pressures on plastome coding regions of C. hypocistis

The establishment of obligate parasitism could trigger the relaxation of selective constraints on chloroplast coding regions that are not functionally important any more. We conducted a formal test that estimates whether selection intensity on plastid coding regions differs in Cytinus compared with its autotrophic relatives. According to our results, all but five protein coding regions present in Cytinus are likely evolving under purifying selection, with selective intensity similar to that in its autotrophic relatives (Table 2). These regions may thus remain functionally important in all the lineages studied here. Similarly, the remaining genes of the highly divergent Epipogium orchids do not show any signs of relaxed selection (Schelkunov et al., 2015). On the other hand, relaxed negative selection was inferred in the Cytinus branch for rpl16, rps3, rps11 and rps19, which implies that the importance of these regions has been reduced without fitness consequences in the holoparasite.

For the region rpl22, high intensification of selective pressure was inferred in the Cytinus branch, with 6% of sites evolving under strong positive selection for Cytinus, and the remaining sites evolving neutrally (44%) or under strong negative selection (50%). In the other branches, half of the sites were inferred to be completely conserved and the other half to be evolving neutrally. Given the small number of positions inferred to be under positive selection and the small size of the gene, we cannot fully discard the idea that this region is diverging in Cytinus just as a result of plastome degeneration, since the region rpl22 is missing or pseudogenized in some parasites (Table S1). It is, however, known that plant lineages can differ in plastid gene need, and indeed, rpl22 is essential for tobacco (Fleischmann et al., 2011) whereas it is not present in the
legume plastome (Doyle et al., 1995). On the other hand, a plausible explanation for positive selection in this lineage would be that this gene is linked somehow to the host–parasite evolutionary arms race (Haraguchi and Sasaki, 1996). Plastid housekeeping genes are not known to be specifically connected to host–parasite interaction, but chloroplast genomes contain genes that are related to key metabolic roles (Barbrook et al., 2006; Krause, 2008), and many aspects of parasite physiology and development could be key to attack hosts. Positive selection has also been suggested in strictly heterotrophic Corallorhiza orchids (Barrett et al., 2014) for atp genes, which are known to be involved in photosynthesis but have been suggested to play additional (and yet unknown) roles as they are highly preserved in some holoparasites (Wicke et al., 2013).

PAML site models consistently show that C. hypocistis has much faster rates of molecular evolution than other Malvales, with dS and dN values at least one order of magnitude larger (Table S5). This result is strongly congruent with previous studies that showed that parasitic plants have raised rates of molecular evolution compared with their close non-parasitic relatives (Nickrent and Starr, 1994; Bromham et al., 2013). Higher mutation rates can be derived from small effective population size (Ebert, 1998; Cutter and Payseur, 2013). This is a plausible explanation for this species, which always occurs in very sparse populations (De Vega et al., 2008) reported low levels of historical gene flow in C. hypocistis populations of the western Mediterranean Basin, with distinct genetic races associated with different hosts. However, we cannot discard the possibility that raised rates are only due to changing selective pressures in the plastid genome; to verify this, nuclear and/or mitochondrial data would be required, as small population size should affect all genomes equally.

**Plastome evolution in parasitic plants**

Highly diverged plastomes are found in all non-photosynthetic species sequenced to date, with the exception of holomycotroph Corallorhiza orchids (Barrett et al., 2014) and the holomycotroph liverwort Anemia mirabilis (Wickett et al., 2008), which show only a slight reduction compared with autotrophic relatives and few pseudogenes. The most common trend among parasites is that plastome reduction tends to increase with host dependence: within the obligate hemiparasite species of Cuscuta, the species with lower photosynthetic activity have smaller plastomes (with fewer genes and smaller intragenic regions) than those with higher photosynthesis (Funk et al., 2007; McNeal et al., 2007); in Santalaceae, obligate parasites of Viscum have lost more functional genes than the facultative parasite Oxyris (Petersen et al., 2015); and in Orobancheae, a strong correlation between gene loss or pseudogenization and degree of host dependence has been reported (Wicke et al., 2013). Additionally, most photosynthetic genes are either absent or have become pseudogenes in non-green parasites.

Only a subset of housekeeping genes are commonly found in parasites and fully mycoheterotrophic plants sequenced to date, which includes six genes encoding ribosomal protein subunits (rpl2, rpl16, rpl36, rps8, rps11, rps14), three tRNAs (trnL-CAU, trnE-UUC, trnM-CAU) and the usual four rRNAs (Table S1). The extremely reduced Pilostyles plastomes differ from this subset, as their genomes only include the genes accD, rps3 and rps4, plus two rRNAs (Bellot and Renner, 2016). The number and type of genes lost seem to be idiosyncratic, but a general path of plastome evolution for parasites has been proposed (Barrett and Davis, 2012; Barrett et al., 2014). According to this model, plastid genome degradation in heterotrophic plants would follow four main stages: (1) degradation of the ndh complex; (2) degradation in ndh and other photosynthesis-related genes; (3) degeneration of all plastid gene systems, including housekeeping genes; and (4) complete or nearly complete loss of the plastid genome (as has been suggested for Rafflesia lagascae; Molina et al., 2014). Following this model, Cytinus is somewhere between stages 3 and 4 (Fig. 1). As stated by Barrett et al. (2014), some clear deviations from this model are found in Cuscuta. Additionally, this model does not explain why positive selection has been detected for some genes in certain parasites. Anyway, the fact that the majority of studies have included one or few parasite species per lineage (and that not all lineages have been studied) still hampers a broader understanding of the mechanisms and pace of plastome reduction.

Concerning the fate of vanished plastid genes, a recent study on seven non-green Orobancheaeae (Cusimano and Wicke, 2015) identified nuclear and mitochondrial insertions for most of the genes lost in the plastome (some corresponding to full copies, others just to fragments), likely as a result of multiple intracellular transfers accumulated over time. However, the majority of these insertions showed high sequence drift and many indels, indicating that they might not be functional.

**Phylogenetic implications**

Our phylogenomic analyses based on coding regions confirm the placement of Cytinaceae within Malvales with strong support, with Malvaceae and Thymelaeaceae as the closest relatives to Cytinaceae. We included here only two Malvales families apart from Cytinaceae because they were the only ones to have complete plastomes (or with a high number of coding regions) available in GenBank. Our results suggest that Malvaceae and Thymelaeaceae are more closely related to each other than to Cytinaceae, a new result not reported before (some interfamilial relationships of Malvales are still unclear today); however, a broader taxon sampling at the family level is needed to confirm this point. Regarding higher taxonomic scale relationships, we here confirm interordinal relationships reported for Malvidae by Ruhfel et al. (2014).

**Future research perspectives in parasitic plastome evolution**

Recent years have seen an upsurge of sequenced plastomes of plant parasites and mycoheterotrophs. However, to really understand the genomic consequences of niche shifts from autotrophy to heterotrophy in plants, a much broader taxon sampling is needed, including all trophic levels of major plant parasite lineages. The information derived from an extensive sampling would likely illuminate how genetic changes in parasites are correlated with the degree of host dependence, time since trophic shift, and gene function. Wider knowledge of the selective pressures on plastome genes of parasites compared
with their autotrophic relatives would also help to unravel whether evolution to parasitism happened stepwise or in a more continuous way; for instance, rapid deletion of large gene blocks has been suggested in Cuscuta, whereas in Orobancheae slow plastome degradation appears to be more likely (Cusimano and Wicke, 2015).

Another interesting research area that remains to be explored more deeply is horizontal gene transfer (HGT) between parasitic plants and their hosts, which is expected to occur thanks to direct physical contact. Indeed, HGT has already been identified in some cases for nuclear and mitochondrial genomes between parasites and host in both directions (e.g. Mower et al., 2004; Davis et al., 2005; Xi et al., 2013), but no plastid-to-plastid transfer has yet been reported to our knowledge. Broader research with genomic and transcriptomic approaches in this area is necessary to unravel the evolutionary consequences derived from HGT and also to identify the mechanisms that facilitate it. Genomic and transcriptomic research could also help to resolve other open questions, such as why plastome genomes have been retained in plant parasites and whether (and which) missing plastid genes have been integrated in the nuclear or mitochondrial genomes.

Concluding remarks

We provide here the complete plastid genome sequence of C. hypocistis, the first parasite genome sequenced for Malvales. This genome is extremely diverged and harbours a spectacular reduction in size and gene number: it mostly encodes ribosomal proteins, tRNAs and rRNAs, and presents no IR regions. We detected very high substitution rates for all coding regions of Cytinus compared with its non-parasitic relatives, which might be associated with small effective population size and/or changing selective pressures. As expected, some regions were inferred to be under relaxed negative selection. On the other hand, strong positive selection was inferred for some sites of the region rpl22 in the Cytinus lineage, which could be explained by a role in the host–parasite evolutionary arms race. However, given the small size of this region, we cannot fully discard the possibility that the divergence found is only the result of plastome degeneration. Phylogenomic analysis of all coding regions confirmed the position of Cytinus within the order Malvales. Future research should incorporate a broad taxon sampling of parasite lineages, including species with different trophic levels, in order to enhance our understanding of mechanisms driving plastome evolution.

SUPPLEMENTARY DATA

Supplementary data are available online at www.aob.oxfordjournals.org and consist of the following. Supplementary Methods: Evolutionary codon models. Figure S1: chloroplast gene phylogeny of Malvidae based on a reduced DNA matrix (only including coding genes present in Cytinus); numbers associated with branches indicate bootstrap support values obtained with 1000 replicates; collapsed monophyletic clades correspond to five Oenothera, 20 Gossypium and 31 Eucalyptus species. Table S1: comparison of the gene content of Cytinus hypocistis with parasitic plants and fully mycoheterotrophic plants sequenced to date. Table S2: list of species included in the phylogenetic analysis and the corresponding GenBank accession numbers. Table S3: summary of the fitted null and alternative codon-substitution models implemented in PAML. Table S4: Akaike information criterion (AIC) scores obtained for site and branch-site models for each putative coding region. The model with the lowest AIC is indicated in bold. Table S5: synonymous (dS) and non-synonymous (dN) substitution rate values of Cytinus versus other Malvales.

ACKNOWLEDGEMENTS

This work received support from the CNRS-INEE (research call in Environmental Genomics ‘APEGE-INEE’, 2012) for the sequencing aspects. The research leading to this paper had received funding from the European Research Council under the European Community’s Seven Framework Programme FP7/2007–2013 Grant Agreement no. 281422 (TEEMBIO).

LITERATURE CITED


Barrett CF, Davis JI. 2012. The plastid genome of the mycoheterotrophic Corallorrhiza striata (Orchidaceae) is in the relatively early stages of degradation. American Journal of Botany 99: 1513–1523.


Bromham L, Bromham L, Cowman PF, Lanfear R. 2013. Parasitic plants have increased rates of molecular evolution across all three genomes. BMC Evolutionary Biology 13: 1–11.


Ruhfel BR, Gitzendanner MA, Soltis PS, Soltis DE, Burleigh JG. 2014. From algae to angiosperms - inferring the phylogeny of green plants (Viridiplantae) from 360 plastid genomes. BMC Evolutionary Biology 14: 23.


Smith DR, Asmail SR. 2014. Next-generation sequencing data suggest that certain nonphotosynthetic green plants have lost their plastid genomes. New Phytologist 204: 7–11.


