

RESEARCH ARTICLE

Open Access

NDH expression marks major transitions in plant evolution and reveals coordinate intracellular gene loss

Tracey A Ruhlman^{1†}, Wan-Jung Chang^{2†}, Jeremy JW Chen^{3†}, Yao-Ting Huang^{4†}, Ming-Tsair Chan^{2†}, Jin Zhang¹, De-Chih Liao², John C Blazier¹, Xiaohua Jin⁵, Ming-Che Shih², Robert K Jansen^{1,6} and Choun-Sea Lin^{2*}

Abstract

Background: Key innovations have facilitated novel niche utilization, such as the movement of the algal predecessors of land plants into terrestrial habitats where drastic fluctuations in light intensity, ultraviolet radiation and water limitation required a number of adaptations. The NDH (NADH dehydrogenase-like) complex of *Viridiplantae* plastids participates in adapting the photosynthetic response to environmental stress, suggesting its involvement in the transition to terrestrial habitats. Although relatively rare, the loss or pseudogenization of plastid NDH genes is widely distributed across diverse lineages of photoautotrophic seed plants and mutants/transgenics lacking NDH function demonstrate little difference from wild type under non-stressed conditions. This study analyzes large transcriptomic and genomic datasets to evaluate the persistence and loss of NDH expression across plants.

Results: Nuclear expression profiles showed accretion of the NDH gene complement at key transitions in land plant evolution, such as the transition to land and at the base of the angiosperm lineage. While detection of transcripts for a selection of non-NDH, photosynthesis related proteins was independent of the state of NDH, coordinate, lineage-specific loss of plastid NDH genes and expression of nuclear-encoded NDH subunits was documented in Pinaceae, gnetophytes, Orchidaceae and Geraniales confirming the independent and complete loss of NDH in these diverse seed plant taxa.

Conclusion: The broad phylogenetic distribution of NDH loss and the subtle phenotypes of mutants suggest that the NDH complex is of limited biological significance in contemporary plants. While NDH activity appears dispensable under favorable conditions, there were likely sufficiently frequent episodes of abiotic stress affecting terrestrial habitats to allow the retention of NDH activity. These findings reveal genetic factors influencing plant/environment interactions in a changing climate through 450 million years of land plant evolution.

Keywords: Streptophyta, NAD(P)H dehydrogenase, Transcriptomics, Cyclic electron flow, Sig4

Background

Key innovations have facilitated novel niche utilization, such as the movement of the algal predecessors of land plants into terrestrial habitats [1] where fluctuations in light intensity, ultraviolet radiation and water limitation required a number of adaptations [1,2]. Among them, mechanisms that permitted more refined control of the

light reactions of photosynthesis may have evolved. The NDH (NADH dehydrogenase-like) complex of plant and algal plastids, which participates in cyclic electron flow (CEF), was initially identified through its homology to the mitochondrial respiratory complex I [3], then shown to be more similar to the NDH-1 pathway of extant cyanobacteria [4,5].

During the light reactions of photosynthesis in cyanobacteria, algae and plants, photons excite pigment/chlorophyll molecules in photosystem II (PSII) at the stromal face of the chloroplast thylakoid membrane. This light energy, in the form of water derived electrons, is

* Correspondence: cslin99@gate.sinica.edu.tw

†Equal contributors

²Agricultural Biotechnology Research Center of Academia Sinica, Agricultural Technology Building, No. 128, Sec. 2, Academia Road, Nankang, Taipei 115, Taiwan

Full list of author information is available at the end of the article

ultimately relayed to photosystem I (PSI) via the intermediate carriers of the electron transport chain to drive proton translocation across the membrane and into the thylakoid lumen. The resulting proton gradient is coupled to ATP synthesis via ATP Synthase with the concomitant reduction of NADP to yield NADPH by ferredoxin-NADP reductase. This process is referred to as linear electron flow and/or transport (LEF). While LEF is the primary pathway for conversion of photons into storable energy, CEF occurs exclusively around photosystem I and contributes to generation of the proton gradient. However the electrons that reduce the plastoquinone (PQ) in CEF are recycled directly from ferredoxin (Fd) allowing for the generation of ATP without the production of NADPH. The resulting high pH differential across the thylakoid membrane induces non-photochemical quenching (NPQ) allowing the dissipation of excess electrons under potentially unfavorable growth conditions including fluctuating, high intensity light, low CO₂ concentration or drought stress [6,7].

Two independent pathways of CEF have been characterized across Streptophyta, the lineage that includes charophyte algae and land plants [6,8]. The genomes of extant cyanobacteria encode the core components of the NDH complex and a pathway for CEF that is sensitive to antimycin A (AA) [9] as has been elucidated in land plants [6]. Sequencing of the *Klebsormidium flaccidum* genome, a streptophyte alga, revealed genes encoding both a functional NDH, with subunit genes in the plastid and nuclear genomes, and components of the nuclear encoded PGR5-dependent (Proton Gradient Regulation5; AA sensitive) pathway of CEF [1].

While both pathways clearly originated in the ancestor of plastids, and both rely on Fd-dependent reduction of PQ, the PGR5-dependent pathway is the main contributor to the pH differential and ultimately ATP generation in CEF [6]. This may be the reason for its ubiquity among the photosynthetic organisms studied to date, whereas NDH CEF has been lost in all photosynthetic lineages examined other than Streptophyta [1,8]. The two pathways are encoded by non-overlapping gene sets and are therefore predictably mechanistically distinct [10].

Cyclic electron flow is essential to efficient photosynthesis and complete inhibition of CEF severely affects LEF *in vivo* [10]; however, the specific role of NDH in CEF remains obscure. Studies support NDH involvement in redox balancing under abiotic stress as it appears to mediate electron flow from stromal reductants to PQ [11-13]. Phenotypes, such as photoinhibition, reduced growth rate and the loss of NPQ induction, are severe where both CEF pathways are restricted [10-14], pronounced when the PGR5-dependent pathway is impaired and subtle when only NDH expression is ablated [5,10,15].

Under favorable growth conditions disruption of NDH has little effect. However CO₂ limitation, extended exposure to low temperatures and high or low light intensities revealed mildly deleterious phenotypes. *Oryza sativa crr6* (required for NDH complex biogenesis) mutants lacked the post illumination increase in chlorophyll fluorescence that is a hallmark of Fd-dependent PQ reduction, and exhibited mild inhibition of all photosynthetic parameters measured and a diminution in plant biomass when grown at 20°C [16]. The disruption of plastid *ndhB* in *Marchantia polymorpha* resulted in a PQ pool that was significantly more reduced at low light intensities relative to the wild type [12]. *Nicotiana tabacum* inactivated for plastid encoded NDH-B showed a reduction in transient chlorophyll fluorescence following actinic illumination, but otherwise performed normally [5]. Repeated, brief exposure to strong light, however, resulted in photoinhibition of PSII and irreversible chlorosis in the same transformed line whereas wild type leaves exposed to the same treatments recovered [13].

Like many plastid-localized, multi-subunit complexes, both plastid- and nuclear-encoded proteins assemble to form the NDH complex [17]. Although relatively rare, the loss or pseudogenization of plastid NDH genes is nonetheless noted across diverse lineages of photoautotrophic seed plants with examples found among gymnosperms [18,19], and both monocot [20-24] and eudicot lineages of angiosperms [25]. As in the preceding experimental examples, these species appear unaffected by the lack of plastid NDH gene expression and several authors have suggested that the missing NDH constituents may have been functionally transferred to the nucleus [18-22,25,26].

Based on phenotypes of mutants/transgenics, NDH function appears to be dispensable under favorable growth conditions; however, the plastid-encoded genes for NDH subunits are conserved across the phylogeny of Streptophyta suggesting a strong selective advantage in retention of NDH function [1]. This study uses large-scale transcriptomic and genomic analyses to evaluate the persistence and loss of NDH expression. Nuclear expression profiles showed that there have been acquisitions of novel NDH genes at key transitions throughout 450 million years of land plant evolution. In the clades where the NDH genes are missing from the plastid genome, no evidence of intracellular horizontal gene transfer of NDH genes was detected. Rather coordinate, lineage-specific loss of expression of nuclear-encoded NDH subunits confirmed the independent loss of NDH in select taxa across the seed plant phylogeny.

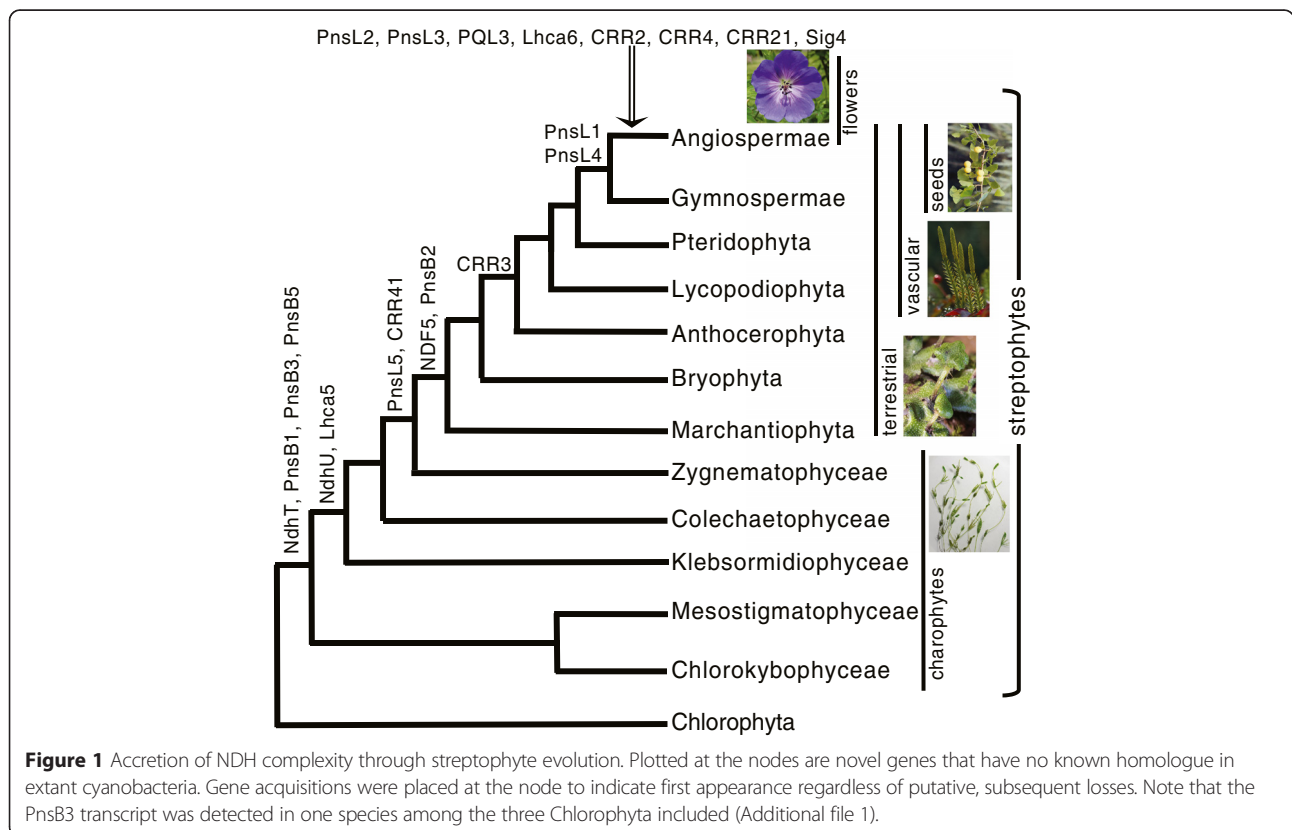
Results and discussion

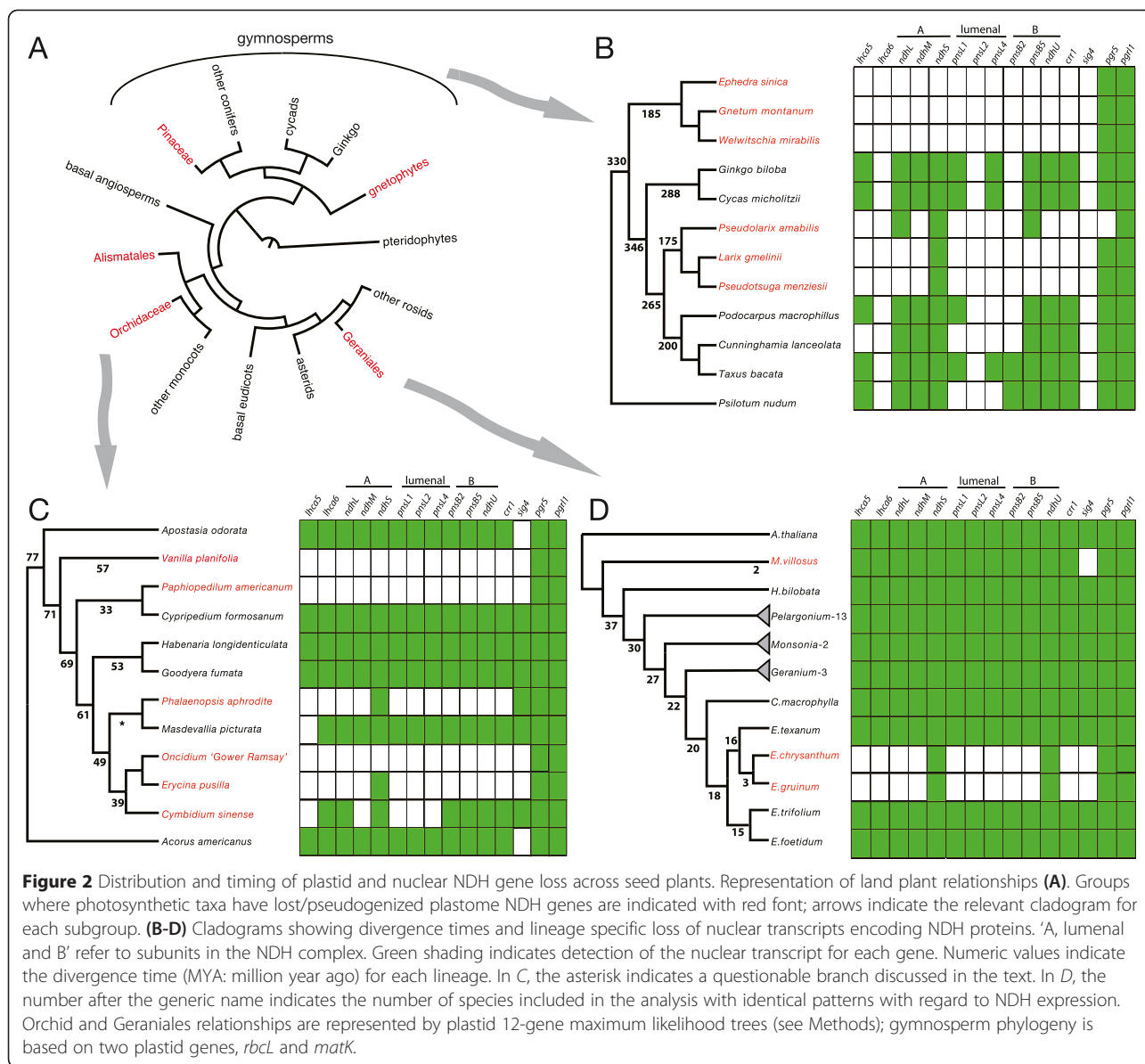
To evaluate the distribution and timing of changes in the NDH gene complement across land plants a subject database comprising nuclear transcriptomes of photoautotrophic Streptophyta species (listed in Additional file 1) was

queried with *Arabidopsis* NDH-related coding sequences and the results of the survey mapped in a phylogenetic context. Queries included constituents of each of the four major subunits containing nuclear proteins, the Lhca (light harvesting complex associated) proteins involved in tethering NDH to PSI during supercomplex formation, assembly and accessory proteins, *ndhF* transcription factor Sig4 and proteins involved in maturation and editing of NDH transcripts in plastids (see Additional file 1 for a complete list). The core constituents of NDH are encoded by genes derived from the common cyanobacterial ancestor of all photosynthetic eukaryotes, 11 genes encoded in the plastid (*ndhA-K*) and five in the nucleus (*ndhL-S*) [17]. Additional related sequences (subunits and auxiliary proteins) were acquired through time and demonstrate an accretion in complexity within NDH, notably at the split between the chlorophyte algae and streptophytes, between charophytes and terrestrial plants, at the origin of seed plants and the base of angiosperms (Figure 1). The more recent gene acquisitions portray an increasing requirement for coordinated control of plastid-encoded gene expression. The majority of gains occurred prior to the diversification of angiosperms and include *ndhF* transcription factor Sig4, and PPR proteins involved in processing of plastid NDH polycistrons and RNA editing (Figure 1; Additional file 1). The co-emergence of Lhca6 at the base of angiosperms could be related to the apparent

requirement for tightly controlled expression of subunits. Supercomplex formation with PSI is thought to positively influence stability of NDH in flowering plants, although a role in NDH assembly has not been discounted [27]. Nuclear control of plastid gene expression may be a means to regulate the stoichiometry of PSI:NDH subunits for optimal efficiency. Whole genome duplication, which occurred in virtually all angiosperms [28], may have been important in generating nuclear substrate sequences for sub- or neo-functionalization yielding the more complex NDH system. Greater control and efficiency of the CEF system, among many other innovations (i.e. [29-31]), may have played a role in the eventual angiosperm radiation into virtually every terrestrial ecosystem.

Despite near ubiquity, a number of exceptional cases exist among photosynthetic species where the NDH genes have been pseudogenized or are entirely missing from plastid genomes (Figure 2A). Among plastomes of Viridiplantae, the loss of NDH genes has been reported for all species of Gnetophyta and Pinaceae, several species in the Orchidaceae and in the long-branch clade within the genus *Erodium* (Geraniaceae). It appears that another species in the Geraniales, *Melianthus villosus*, is undergoing NDH loss as four of its plastome-encoded NDH genes are pseudogenes [32]. Plastid NDH gene loss was recently revealed in the monocot family Hydrocharitaceae, which harbors at least three species of aquatic





angiosperms that have undergone NDH disruption (*Najas Vallisneria*, *Thalassia*) with two further losses likely among the marine 'seagrasses' in the order Alismatales (*Posidonia*, *Amphibolis*) [23,24].

Of the non-NDH plastome genes that have been lost or pseudogenized, several have been detected in the nuclear genome where their expression is regulated by nuclear factors and subsequent products targeted back to plastids. Although several groups have suggested that plastid NDH pseudogenes may be functionally represented by nuclear sequences [18-22,25], no reports describe a functionally transferred, nuclear NDH gene of recent plastid origin. A number of nuclear-encoded NDH subunits and accessory proteins have been elucidated [6,8], which are targeted to the plastid where they

assemble with plastome-encoded subunits at the thylakoid membrane. Figures 2B-D present an abbreviated graphical overview of the pattern of NDH nuclear gene loss. In all investigated lineages where the plastid sequences have been lost, no transcripts were detected that could represent recently transferred plastid NDH genes. Furthermore, transcripts reporting expression of nuclear-encoded NDH subunits are, to varying degrees, absent (Figure 2B-D, Additional file 1). To confirm that the lack of nuclear NDH transcripts is exclusive to NDH genes and not a more general phenomenon and to demonstrate that there were sufficient data to detect such transcripts were they present, transcriptomes of plastid NDH-deleted species and those that retain NDH in the plastome were queried for the presence of transcripts

encoding members of gene families to which NDH subunits belong. Regardless of the state of the NDH system nearly all searched non-NDH transcripts inherent to photosynthesis and plastid respiration were detected (Additional file 2). In a very few cases where data were mined from public databases, expected transcripts (i.e. PGR5) were not detected. Overall the mined data yielded findings highly similar to those found using the data produced for this study, the collection of which was standardized according to Zhang et al. [33]. The downloaded transcriptomes varied broadly in terms of tissue type, library construction, sequence length and read depth, however they were of sufficiently high quality to detect most if not all genes searched for this study. The use of *Arabidopsis* reference sequences may have also allowed some transcripts to go undetected due to high levels of divergence between the reference and the taxa examined.

Indeed, the 'patchy' signal detected for *lhca5* (Additional file 1) prior to its consistent presence in angiosperms could be a result of the caveats described above. For this study the parameters used to define a transcript as present or absent, as described in the methods, were set *a priori*, as was the decision to report the first incidence of detection regardless of subsequent lack of detection in more recently diverged lineages. The chlorophyte algae *Chlamydomonas reinhardtii* and *Ostreococcus tauri*, among others, are thought to contain a gene encoding *Lhca5* (GenBank, ABD37907 and AAY27547.1), however the transcript was not detected using the present data mining strategy. Of course these species are known to lack NDH and likely their *lhca5* sequences are not similar enough to be detected using *Arabidopsis* queries. Even in more closely related groups, i.e. gymnosperms and angiosperms, there is a margin for error although calls of presence or absence are expected to be more reliable between more recently diverged groups. The *lhca5* transcript was detected in seven of 11 gymnosperm species that are competent for NDH CEF (Additional file 1).

The presence of the AA-sensitive pathway described by Arnon et al. [34], the predominant pathway for CEF in plants, was inferred by the identification of nuclear transcripts encoding PGR5 and PGRL1 (A/B) in nearly all species surveyed (Figure 2B-D; Additional file 1). Although wholesale loss of nuclear-encoded NDH gene expression is typical of the taxa in which plastome NDH genes are absent, there is some variation across this select group. One ancestral NDH subcomplex A gene that is retained across the majority of taxa except gnetophytes is the nuclear gene encoding *NdhS* (Crr31), responsible for high affinity binding of ferredoxin [11] (Figure 2; Additional file 1). In experiments where this function is ablated, NDH complex formation is concomitantly lost. However *NdhS* accumulated in the

thylakoid membranes of mutants that do not express NDH suggesting this protein may have an additional function in plastids [11].

The gene encoding the thylakoid immunophilin *PnsL5* (CYP20-2) lacks a cyanobacterial homolog and is nearly ubiquitous across the species surveyed, apparently arising at the node leading to Zygnematomphyceae, the proposed sister lineage to land plants (Figure 1, Additional file 1). Although *PnsL5* has been identified in NDH and NDH-PSI [35] preparations as a luminal subcomplex protein and its expression was strongly reduced in NDH mutants with defects in the hydrophobic domains, mutants lacking *PnsL5* expression were not affected in complex stability or NDH activity [36]. There remains some question regarding its status as a bona fide subunit of NDH; while *PnsL5/CYP20-2* may indeed be an NDH-associated protein this isomerase has demonstrated roles in gibberilic acid and brassinosteroid signaling in *Triticum* and *Arabidopsis*, respectively [37,38]. Given that phytohormone signaling has been proposed as an adaptation facilitating colonization of land by plants [30,39], the emergence of *PnsL5/CYP20-2* in the sister lineage to land plants, tempts speculation of a dual role for this protein in the early adaptation to land.

A phylogenetic context permits evaluation of the distribution of retention/loss as a function of evolutionary time and thus the estimation of the number of independent losses that have occurred. For example within gymnosperms there are two distinct groups lacking plastid NDH genes; all sampled gnetophytes and Pinaceae. There has been considerable controversy regarding the phylogenetic position of gnetophytes among gymnosperms with several different placements supported by different data sets (i.e. gnepine, gnetifer, anthopyte, gnetophytes sister to all other gymnosperms) [40-45]. Accordingly, interpretation of the pattern of NDH gene loss will be affected by which of these hypotheses is correct. The tree shown in Figure 2B is based on limited taxon sampling and two plastid genes, *rbcl* and *matK*, and reflects the relationships identified in Lee et al. [46], placing gnetophytes as sister to all other gymnosperms. The complete excision of all plastid NDH genes in *Gnetum* and *Ephedra* and a single detectable pseudogene in *Welschitchia* [47,48] suggest that the loss in Pinaceae may be more recent as these plastomes retain many more NDH pseudogenes [18]. If indeed the relationships inferred by Lee et al. are correct, then two independent losses would be appropriately assigned. Alternatively, if gnetophytes are sister to Pinaceae, then the most parsimonious interpretation is of a single loss. Nonetheless, these events are relatively ancient (Figure 2B) and both groups have similarly lost function of all plastid genes for NDH along with nuclear NDH gene expression with the exception of *PnsL5* and, at least for the Pinaceae species, *ndhS* (Additional file 1). As seen in Orchidaceae,

however, events interpreted as synapomorphic may in fact represent independent events whose signals have been obscured through evolutionary time. This phenomenon is well illustrated by the repeated, independent loss of the plastome *infA* and *accD* genes among others [49].

Among the Orchidaceae four independent losses of NDH are indicated (Figure 2C). However the 12-gene phylogeny presented here is limited to the 11 species for which both plastome and transcriptome data are available and differs somewhat in topology from trees incorporating hundreds of species. Expanded phylogenies have placed *Masdevallia* outside the clade of higher epidendroids that it groups with here [50]. This alternative placement predicts three independent NDH losses among orchids. Contrary to the suggestion that Sig4 evolved after the diversification of eudicots [51], *sig4* transcripts were detected among the orchid species that retain NDH function as well as in *Amborella* (Additional file 1).

Recent loss of plastid NDH genes is observed in the long-branch clade (LBC) of *Erodium*. Losses have been confirmed in three among 13 species through complete plastome sequencing; losses in all remaining species were inferred through PCR survey for the *ndhD* gene [25] suggesting that the loss occurred prior to the diversification of the clade (3 MYA). Within the Geraniales there appears to be an even more recent loss in the genus *Melianthus* where four NDH pseudogenes have been identified in the plastome [32]. This clade is believed to have diversified approximately 2 MYA and retains some apparently translatable plastome NDH sequences. Probing the *M. villosus* transcriptome revealed that while most NDH related transcripts were detected, the *ndhF* transcription factor Sig4 is transcribed as a pseudogene (Figure 2D).

Conclusions

These data allow insight into the distribution and timing of NDH gene loss across seed plants and suggest at least limited dispensability of NDH function. In no case were functional nuclear copies of plastome NDH sequences identified but rather the degradation and loss of nuclear-encoded interacting proteins was revealed. The broad phylogenetic distribution of NDH loss and the subtle phenotypes of mutants suggest a high propensity for gene loss and may be indicative of its limited biological significance in contemporary plants [52]. On an evolutionary time scale, however, climate fluctuates in virtually all habitats such that NDH function may be intermittently significant. While NDH activity appears dispensable under favorable conditions there may be sufficiently frequent episodes of abiotic stress affecting terrestrial habitats to allow the retention of NDH activity throughout land plant evolution. In fact, the accumulation of RNA editing sites in plastid NDH genes has been interpreted as

arising from such recurrent “dispensability and rescue” of NDH activity [8].

The timing of changes in the NDH gene complement supports the hypothesis that the NDH system, among other adaptations, has been involved in the transition to terrestrial habitats and possibly other key innovations throughout 450 million years of land plant evolution. Given the present outlook on climate change a deeper understanding of the genetic factors influencing the adaptation of land plants to novel conditions will inform management programs.

Methods

Plant materials

Tissue samples for species of Geraniales were harvested from the living collection at University of Texas at Austin. All specimens have been vouchered and deposited in the UT Plant Resources Center (TEX-LL). Voucher numbers are listed in Additional file 3.

Apostasia odorata was collected from Yunnan, China. *Cypripedium formosanum* was obtained from highland experimental farm, National Taiwan University in Meifeng, Taiwan. Three commercial Orchidaceae species were purchased from a local grower in Taiwan, two Epidendroideae (*Masdevallia picturata* and *Erycina pusilla*) and one Orchidoideae (*Goodyera fumata*). These three species were grown in the greenhouses at Academia Sinica, Taipei, and National Chung Hsing University, Taichung, Taiwan.

RNA isolation, transcriptome sequencing and assembly

Total RNA from newly emerged leaves of 26 species in Geraniales and four tissues (emergent and expanded leaves, roots and flowers) of *Pelargonium x hortorum* was isolated and cDNA libraries constructed following the protocols described in Zhang et al. [33]. Transcriptome sequencing was performed on the Illumina HiSeq™ 2000 platform (Illumina, San Diego, CA). For Orchidaceae, total RNA was extracted from young leaves by TRIzol® reagent (Life-Technologies, Taipei City, Taiwan). Agilent 2100 Bioanalyzer (Agilent Technologies, Taipei City, Taiwan) was used to confirm total RNA quality. Six paired-end RNA-Seq libraries (*A. odorata*, *C. formosanum*, *M. picturata*, *G. fumata*, *E. pusilla* and *H. longidenticulata*) were constructed using the Illumina TruSeq™ Stranded mRNA HT, insert size was 300 bp. The libraries were sequenced on Illumina NextSeq™ 500 paired-end system using a NextSeq™ 500 Mid output kit (300 cycles; Illumina). Sequence data of Geraniales and the five species of orchid were preprocessed and assembled as described in Zhang et al. [33]. Transcriptome data for the remaining species examined were downloaded from the NCBI Sequence Read Archive (SRA, Additional file 1). Reads were assembled with Trinity release 2013/11/10 (

sourceforge.net/projects/trinityrnaseq/) using the parameters “-JM 100G -full_cleanup -min_contig_length 200 -CPU 24” on a 24-core 3.33 GHz linux work station with 1 TB memory at the Texas Advanced Computing Center (TACC, <http://www.tacc.utexas.edu/>).

Identification of nuclear-encoded *ndh* genes

The protein sequences of reference genes of *Arabidopsis* were downloaded from TAIR [53]; accession numbers are given in Additional files 1 and 2. Reference gene sequences were aligned to the transcriptome assemblies using TBLASTN with e-value 1e-5 to extract the nuclear encoded candidate genes of the NDH complex (See Additional file 4). The candidate genes were then aligned to the *Arabidopsis* transcriptome using BLASTX with default settings, and the gene was considered present if the top hit was the corresponding reference gene. If the top hit was not the corresponding gene, manual inspection of these candidate genes was performed to resolve potential chimeric assembly problems and the candidate genes were confirmed again with BLASTX. All newly generated sequences have been submitted to NCBI GenBank, accession numbers are in Additional file 4.

Phylogeny construction

The phylogenies of 11 species of orchids plus *Acorus* and 26 species of Geraniales plus *Arabidopsis* were generated by RAxML with parameters “raxmlHPC-PTHREADS-SSE3 -f a -x 12345 -p 12345 -T 12 -m GTRGAMMAI -N 100” using 12 plastid genes (*atpA*, *atpB*, *atpI*, *ccsA*, *cemA*, *matK*, *petA*, *rbcL*, *rpoB*, *rpoC1*, *rpoC2*, *rps2*), and bootstrap values were generated using RAxML with 100 replicates and the above settings. The gymnosperm phylogeny of 11 species plus *Psilotum* was generated by the same parameters except that two sequences, *matK* and *rbcL* were employed. For Figure 1 and 2A, trees were drawn manually based on relationships depicted in [54] and [45,46].

Divergence time estimation

Divergence time estimates were derived from previous studies of gymnosperms [55,56], Orchidaceae [50] and Geraniales [57,58].

Availability of supporting data

Phylogenetic datasets are available in Dryad Digital Repository (<http://dx.doi.org/10.5061/dryad.h9m07>). Sequence data are available in the Short Read Archive at GenBank, <http://www.ncbi.nlm.nih.gov/sra>; accession numbers are listed in Additional files 1, 2, 4.

Additional files

Additional file 1: Expanded survey of nuclear encoded NDH transcripts. This table includes data for 106 species, both generated in this study and downloaded from a number of sources.

Additional file 2: Detection *psbP*, *psbQ*, *lhca*, FKBP and PGR gene family transcripts in gymnosperms, Orchidaceae and Geraniales. This table includes data for 50 species as indicated in the title and reports the expression of nuclear genes unrelated to NDH.

Additional file 3: Voucher information for 26 species of Geraniales. University of Texas Plant Resources Center voucher numbers for Geraniales specimens used in this study.

Additional file 4: Accession numbers for all specific transcript sequences generated in this study. Collection of GenBank accession numbers for all new plastid genome and individual nuclear transcript sequences reported in this study.

Abbreviations

LEF: Linear electron flow; CEF: Cyclic electron flow; PSI: Photosystem I; PSII: Photosystem II; AA: Antimycin A; NDH: Plastid NAD(P)H dehydrogenase-like complex; MYA: Million years ago; Lhca: Light harvesting complex genes/proteins; LBC: Long branch clade (as in *Erodium*).

Competing interests

The authors declare that they have no competing interests.

Authors' contributions

TAR prepared RNA samples for Geraniales, designed analyses, assisted in figure and table production and wrote the manuscript. WJC, JJWC, YTH, MTC and CSL performed orchid sequencing, assembly and annotation, and performed Blast searches of public sequences for genes of interest. WJC, DCL and CSL contributed to figure and table production. DCL, XJ and MCS isolated orchid RNA. DCL participated in orchid bioinformatics analyses. JZ assembled and annotated transcriptomes for all Geraniales and several orchids, and generated phylogenies for Figures 2B-D. RKJ designed analyses, assisted in figure and table production. JCB provided sequences for 12-gene phylogenies (Figure 1D). CSL developed initial experimental strategy and designed analyses. All authors read and approved the final manuscript.

Acknowledgements

Support was provided by the National Science Foundation (IOS-1027259 to RJK and TAR) and by Innovative Translational Agricultural Research Administrative Office (to CSL). The authors thank Der-Ming Yeh, Li-Ya Chao and Jin-Jue Yue for providing orchid materials and analysis, Robin Parer at www.geraniaceae.com for Geraniaceae materials, and Don Levin, Tom Juenger, Mike Ryan and two anonymous reviewers for comments on an earlier version of the manuscript.

Author details

¹Department of Integrative Biology, University of Texas at Austin, Austin, TX, USA. ²Agricultural Biotechnology Research Center of Academia Sinica, Agricultural Technology Building, No. 128, Sec. 2, Academia Road, Nankang, Taipei 115, Taiwan. ³Institute of Biomedical Sciences, National Chung Hsing University, Taichung, Taiwan. ⁴Department of Computer Science and Information Engineering, National Chung Cheng University, Chia-Yi, Taiwan. ⁵Institute of Botany, Chinese Academy of Sciences, Beijing 100093, China. ⁶Department of Biological Science, Biotechnology Research Group, King Abdulaziz University, Jeddah 21589, Saudi Arabia.

Received: 1 December 2014 Accepted: 30 March 2015

Published online: 11 April 2015

References

- Hori K, Maruyama F, Fujisawa T, Togashi T, Yamamoto N, Seo M, et al. *Klebsormidium flaccidum* genome reveals primary factors for plant terrestrial adaptation. *Nat Commun*. 2014;5:3978.
- Rensing SA, Lang D, Zimmer AD, Terry A, Salamov A, Shapiro H, et al. The *Physcomitrella* genome reveals evolutionary insights into the conquest of land by plants. *Science*. 2008;319:64–9.

3. Matsubayashi T, Wakasugi T, Shinozaki K, Yamaguchi-Shinozaki K, Zaita N, Hidaka T, et al. Six chloroplast genes (*ndhA-F*) homologous to human mitochondrial genes encoding components of the respiratory chain NADH dehydrogenase are actively expressed: determination of the splice sites in *ndhA* and *ndhB* pre-mRNAs. *Mol Gen Genet*. 1987;210:385–93.
4. Friedrich T, Steinmüller K, Weiss H. The proton-pumping respiratory complex I of bacteria and mitochondria and its homologue in chloroplasts. *FEBS Lett*. 1995;367:107–11.
5. Shikanai T, Endo T, Hashimoto T, Yamada Y, Asada K, Yokota A. Directed disruption of the tobacco *ndhB* gene impairs cyclic electron flow around photosystem I. *Proc Natl Acad Sci U S A*. 1998;95:9705–9.
6. Shikanai T. Central role of cyclic electron transport around photosystem I in the regulation of photosynthesis. *Curr Opin Biotechnol*. 2014;26:25–30.
7. Buchanan B, Gruissem W, Jones R. *Biochemistry & Molecular Biology of Plants*. Hoboken, NJ USA: Wiley; 2015.
8. Martin M, Sabater B. Plastid *ndh* genes in plant evolution. *Plant Physiol Biochem*. 2010;48:636–45.
9. Yeremenko N, Jeanjean R, Prommeenate P, Krasikov V, Nixon PJ, Vermaas WFJ, et al. Open reading frame *ssr2016* is required for antimycin A-sensitive photosystem I-driven cyclic electron flow in the cyanobacterium *Synechocystis* sp. PCC 6803. *Plant Cell Physiol*. 2005;46:1433–6.
10. Munekage Y, Hashimoto M, Miyake C, Tomizawa K-I, Endo T, Tasaka M, et al. Cyclic electron flow around photosystem I is essential for photosynthesis. *Nature*. 2004;429:579–82.
11. Yamamoto H, Peng L, Fukao Y, Shikanai T. An Src homology 3 domain-like fold protein forms a ferredoxin binding site for the chloroplast NADH dehydrogenase-like complex in *Arabidopsis*. *Plant Cell*. 2011;23:1480–93.
12. Ueda M, Kuniyoshi T, Yamamoto H, Sugimoto K, Ishizaki K, Kohchi T, et al. Composition and physiological function of the chloroplast NADH dehydrogenase-like complex in *Marchantia polymorpha*. *Plant J*. 2012;72:683–93.
13. Endo T, Shikanai T, Takabayashi A, Asada K, Sato F. The role of chloroplastic NAD(P)H dehydrogenase in photoprotection. *FEBS Lett*. 1999;457:5–8.
14. Suorsa M, Järvi S, Grieco M, Nurmi M, Pietrzykowska M, Rantala M, et al. PROTON GRADIENT REGULATION5 is essential for proper acclimation of *Arabidopsis* photosystem I to naturally and artificially fluctuating light conditions. *Plant Cell Online*. 2012;24:2934–48.
15. Horváth EM, Peter SO, Joët T, Rumeau D, Courmac L, Horváth GV, et al. Targeted inactivation of the plastid *ndhB* gene in tobacco results in an enhanced sensitivity of photosynthesis to moderate stomatal closure. *Plant Physiol*. 2000;123:1337–50.
16. Yamori W, Sakata N, Suzuki Y, Shikanai T, Makino A. Cyclic electron flow around photosystem I via chloroplast NAD(P)H dehydrogenase (NDH) complex performs a significant physiological role during photosynthesis and plant growth at low temperature in rice. *Plant J*. 2011;68:966–76.
17. Ifuku K, Endo T, Shikanai T, Aro E-M. Structure of the chloroplast NADH dehydrogenase-like complex: Nomenclature for nuclear-encoded subunits. *Plant Cell Physiol*. 2011;52:1560–8.
18. Wakasugi T, Tsudzuki J, Ito S, Nakashima K, Tsudzuki T, Sugiura M. Loss of all *ndh* genes as determined by sequencing the entire chloroplast genome of the black pine *Pinus thunbergii*. *Proc Natl Acad Sci U S A*. 1994;91:9794–8.
19. Braukmann TWA, Kuzmina M, Stefanović S. Loss of all plastid *ndh* genes in Gnetales and conifers: extent and evolutionary significance for the seed plant phylogeny. *Curr Genet*. 2009;55:323–37.
20. Neyland R, Urbatsch LE. Phylogeny of subfamily Epidendroideae (Orchidaceae) inferred from *ndhF* chloroplast gene sequences. *Am J Bot*. 1996;83:1195–206.
21. Chang C-C, Lin H-C, Lin I-P, Chow T-Y, Chen H-H, Chen W-H, et al. The chloroplast genome of *Phalaenopsis aphrodite* (Orchidaceae): Comparative analysis of evolutionary rate with that of grasses and its phylogenetic implications. *Mol Biol Evol*. 2006;23:279–91.
22. Wu F-H, Chan M-T, Liao D-C, Hsu C-T, Lee Y-W, Daniell H, et al. Complete chloroplast genome of *Oncidium Gower Ramsey* and evaluation of molecular markers for identification and breeding in Oncidiinae. *BMC Plant Biol*. 2010;10:68.
23. Iles WJD, Smith SY, Graham SW. A well-supported phylogenetic framework for the monocot order Alismatales reveals multiple losses of the plastid NADH dehydrogenase complex and a strong long-branch effect. In: *Early Events in Monocot Evolution*. Cambridge UK: Cambridge University Press; 2013.
24. Peredo EL, King UM, Les DH. The plastid genome of *Najas flexilis*: Adaptation to submersed environments is accompanied by the complete loss of the NDH complex in an aquatic angiosperm. *PLoS One*. 2013;8:e68591.
25. Blazier JC, Guisinger MM, Jansen RK. Recent loss of plastid-encoded *ndh* genes within *Erodium* (Geraniaceae). *Plant Mol Biol*. 2011;76:263–72.
26. Timmis JN, Ayliffe MA, Huang CY, Martin W. Endosymbiotic gene transfer: Organelle genomes forge eukaryotic chromosomes. *Nat Rev Genet*. 2004;5:123–35.
27. Peng L, Shikanai T. Supercomplex formation with photosystem I is required for the stabilization of the chloroplast NADH dehydrogenase-like complex in *Arabidopsis*. *Plant Physiol*. 2011;155:1629–39.
28. Soltis DE, Bell CD, Kim S, Soltis PS. Origin and early evolution of angiosperms. *Ann NY Acad Sci*. 2008;1133:3–25.
29. Hoson T, Wakabayashi K. Role of the plant cell wall in gravity resistance. *Phytochemistry*. 2014. doi:10.1016/j.phytochem.2014.08.022.
30. Mouradov A, Spangenberg G. Flavonoids: a metabolic network mediating plants adaptation to their real estate. *Front Plant Sci*. 2014;5:620.
31. Edel KH, Kudla J. Increasing complexity and versatility: How the calcium signaling toolkit was shaped during plant land colonization. *Cell Calcium*. 2015;57:231–46.
32. Weng M-L, Blazier JC, Govindu M, Jansen RK. Reconstruction of the ancestral plastid genome in Geraniaceae reveals a correlation between genome rearrangements, repeats, and nucleotide substitution rates. *Mol Biol Evol*. 2014;31:645–59.
33. Zhang J, Ruhlman TA, Mower JP, Jansen RK. Comparative analyses of two Geraniaceae transcriptomes using next-generation sequencing. *BMC Plant Biol*. 2013;13:228.
34. Arnon DI, Allen MB, Whatley FR. Photosynthesis by isolated chloroplasts. *Nature*. 1954;174:394–6.
35. Peng L, Fukao Y, Fujiwara M, Takami T, Shikanai T. Efficient operation of NAD(P)H dehydrogenase requires supercomplex formation with photosystem I via minor LHCI in *Arabidopsis*. *Plant Cell*. 2009;21:3623–40.
36. Sirpiö S, Holmström M, Battchikova N, Aro E-M. AtCYP20-2 is an auxiliary protein of the chloroplast NAD(P)H dehydrogenase complex. *FEBS Lett*. 2009;583:2355–8.
37. Li B, Xu W, Xu Y, Zhang Y, Wang T, Bai Y, et al. Integrative study on proteomics, molecular physiology, and genetics reveals an accumulation of cyclophilin-like protein, TaCYP20-2, leading to an increase of Rht protein and dwarf in a novel GA-insensitive mutant (*gaid*) in wheat. *J Proteome Res*. 2010;9:4242–53.
38. Zhang Y, Li B, Xu Y, Li H, Li S, Zhang D, Mao Z, Guo S, Yang C, Weng Y, Chong K: The cyclophilin CYP20-2 modulates the conformation of BRASSINAZOLE-RESISTANT1, which binds the promoter of FLOWERING LOCUS D to regulate flowering in *Arabidopsis*. *Plant Cell* 2013: tpc.113.110296.
39. Zi J, Mafu S, Peters RJ. To gibberellins and beyond! Surveying the evolution of (di)terpenoid metabolism. *Annu Rev Plant Biol*. 2014;65:259–86.
40. Bowe LM, Coat G, de Pamphilis CW. Phylogeny of seed plants based on all three genomic compartments: Extant gymnosperms are monophyletic and Gnetales' closest relatives are conifers. *Proc Natl Acad Sci U S A*. 2000;97:4092–7.
41. Chaw S-M, Parkinson CL, Cheng Y, Vincent TM, Palmer JD. Seed plant phylogeny inferred from all three plant genomes: Monophyly of extant gymnosperms and origin of Gnetales from conifers. *Proc Natl Acad Sci U S A*. 2000;97:4086–91.
42. Burleigh JG, Mathews S. Phylogenetic signal in nucleotide data from seed plants: implications for resolving the seed plant tree of life. *Am J Bot*. 2004;91:1599–613.
43. Soltis DE, Soltis PS, Zanis MJ. Phylogeny of seed plants based on evidence from eight genes. *Am J Bot*. 2002;89:1670–81.
44. Wickett NJ, Mirarab S, Nguyen N, Warnow T, Carpenter E, Matasci N, Ayyampalayam S, Barker MS, Burleigh JG, Gitzendanner MA, Ruhfel BR, Wafala E, Der JP, Graham SW, Mathews S, Melkonian M, Soltis DE, Soltis PS, Miles NW, Rothfels CJ, Pokorný L, Shaw AJ, DeGironimo L, Stevenson DW, Surek B, Villarreal JC, Roure B, Philippe H, dePamphilis CW, Chen T, et al.: Phylotranscriptomic analysis of the origin and early diversification of land plants. *Proc Natl Acad Sci USA* 2014:201323926.
45. Ruhfel BR, Gitzendanner MA, Soltis PS, Soltis DE, Burleigh JG. From algae to angiosperms—inferring the phylogeny of green plants (Viridiplantae) from 360 plastid genomes. *BMC Evol Biol*. 2014;14:23.

46. Lee EK, Cibrian-Jaramillo A, Kolokotronis S-O, Katari MS, Stamatakis A, Ott M, et al. A functional phylogenomic view of the seed plants. *PLoS Genet*. 2011;7:e1002411.
47. Wu C-S, Lai Y-T, Lin C-P, Wang Y-N, Chaw S-M. Evolution of reduced and compact chloroplast genomes (cpDNAs) in gnetophytes: Selection toward a lower-cost strategy. *Mol Phylogenet Evol*. 2009;52:115–24.
48. McCoy SR, Kuehl JV, Boore JL, Raubeson LA. The complete plastid genome sequence of *Welwitschia mirabilis*: an unusually compact plastome with accelerated divergence rates. *BMC Evol Biol*. 2008;8:130.
49. Ruhlman TA, Jansen RK. The Plastid Genomes of Flowering Plants. In: Maliga P, editor. *Chloroplast Biotechnology: Methods and Protocols*. Volume 1132. New York: Springer Science and Business Media, LLC; 2013.
50. Gustafsson ALS, Verola CF, Antonelli A. Reassessing the temporal evolution of orchids with new fossils and a Bayesian relaxed clock, with implications for the diversification of the rare South American genus *Hoffmannseggella* (Orchidaceae: Epidendroideae). *BMC Evol Biol*. 2010;10:177.
51. Seliverstov AV, Lysenko EA, Lyubetsky VA. Rapid evolution of promoters for the plastome gene *ndhF* in flowering plants. *Russ J Plant Physiol*. 2009;56:838–45.
52. Krylov DM, Wolf YI, Rogozin IB, Koonin EV. Gene loss, protein sequence divergence, gene dispensability, expression level, and interactivity are correlated in eukaryotic evolution. *Genome Res*. 2003;13:2229–35.
53. Lamesch P, Berardini TZ, Li D, Swarbreck D, Wilks C, Sasidharan R, Muller R, Dreher K, Alexander DL, Garcia-Hernandez M, Karthikeyan AS, Lee CH, Nelson WD, Ploetz L, Singh S, Wensel A, Huala E. The Arabidopsis Information Resource (TAIR): improved gene annotation and new tools. *Nucleic Acids Res*. 2011;gkr1090.
54. Leliaert F, Smith DR, Moreau H, Herron MD, Verbruggen H, Delwiche CF, et al. Phylogeny and molecular evolution of the green algae. *Crit Rev Plant Sci*. 2012;31:1–46.
55. Crisp MD, Cook LG. Cenozoic extinctions account for the low diversity of extant gymnosperms compared with angiosperms. *New Phytol*. 2011;192:997–1009.
56. Leaché AD. The Timetree of Life. S. Blair Hedges and Sudhir Kumar, editors. *Integr Comp Biol*. 2010;50:141–2.
57. Palazzesi L, Gottschling M, Barreda V, Weigend M. First Miocene fossils of Vivianiaceae shed new light on phylogeny, divergence times, and historical biogeography of Geraniales. *Biol J Linn Soc*. 2012;107:67–85.
58. Fiz O, Vargas P, Alarcón M, Aedo C, García JL, Aldasoro JJ. Phylogeny and historical biogeography of Geraniaceae in relation to climate changes and pollination ecology. *Syst Bot*. 2008;33:326–42.

Submit your next manuscript to BioMed Central and take full advantage of:

- Convenient online submission
- Thorough peer review
- No space constraints or color figure charges
- Immediate publication on acceptance
- Inclusion in PubMed, CAS, Scopus and Google Scholar
- Research which is freely available for redistribution

Submit your manuscript at
www.biomedcentral.com/submit

