Chloroplast DNA evidence for the origin of the genus *Heterogaura* from a species of *Clarkia* (Onagraceae)

(Chloroplast DNA phylogeny/restriction-site variation/plant taxonomy)

KENNETH J. SYTSMAN* and LESLIE D. GOTTLIEB**

*Department of Genetics, University of California, Davis, CA 95616

*Submitted by Genetics, University of California, July 12, 1986

ABSTRACT Restriction-site variation in chloroplast DNA was examined in the morphologically distinct and monotypic genus *Heterogaura* and the related speciose genus *Clarkia* (Onagraceae), both native to California. Of the 605 restriction sites surveyed, a total of 119 mutations were identified. Of these, 55 were shared by at least two species and were used to construct a most parsimonious phylogenetic tree. This analysis, as well as one based on a distance metric, provided evidence that *Heterogaura* and *Clarkia dudleyana*, a member of a phylogenetically advanced section, share a more recent common ancestor than either does with any other species. The two species are more closely related than nearly all pairs of *Clarkia* tested. The origin of *Heterogaura* from within another genus raises important questions about the adequacy of morphological data and suggests that the relationships of other well-known monotypic plant genera should be reinvestigated.

Many plant taxa have been described that possess unusual and distinctive morphologies, which so sharply delimit them that they are given generic rank. Implicit in such a classification is that the common ancestor of two related genera is more ancient than the common ancestor of any pair of species within either genus. However, dramatic morphological divergence need not imply much genetic divergence (1) and can obscure a close phylogenetic relationship. In this paper, we use restriction enzyme analysis of chloroplast DNA (cpDNA) to examine the relationship of the monotypic genus *Heterogaura* (Onagraceae) to the related and speciose genus *Clarkia* (which includes ~33 diploid and 10 polyploid species).

The monotypic *Heterogaura heterandra* (Torr.) Cov. is a self-pollinating annual plant limited to California and Oregon. Its floral and fruit characters are so unusual that it has been maintained as a genus since 1866. Its flowers possess four fertile and four sterile anthers and a smooth unlodged capitate stigma. Its ovaries contain up to four ovules and mature into round nut-like indehiscent fruits with one or two seeds. In contrast, *Clarkia* flowers have four-lobed stigmas (the self-pollinating species have reduced lobing), generally eight stamens, and an elongated many-seeded capsule, which splits along four sutures to release the seeds. Our analysis shows that *H. heterandra* is closely related to the outcrossing annual *Clarkia dudleyana*, which belongs to sect. *peripetasma*. The two species occupy similar habitats on the western slope of the Sierra Nevada in California and have the same chromosome number (*n* = 9). Sect. *peripetasma* is known with certainty to be phylogenetically advanced within *Clarkia* because it possesses duplicated genes encoding the cytosolic isozyme of phosphoglucone isomerase (2, 3). This result, indicating the origin of a new genus within an advanced section of another genus, is particularly noteworthy because the Onagraceae is one of the best-known plant families (4).

The circular DNA molecule of chloroplasts has been particularly useful in phylogenetic studies of plants. cpDNA is relatively large (140–175 kilobase pairs) and is present in high copy number within leaf cells, permitting its ready isolation (5). Its high degree of sequence conservation facilitates use of heterologous probes to visualize restriction fragments on Southern blots. Comparative mapping of cpDNA among congeneric species has provided detailed phylogenetic relationships in a number of Angiosperm genera (6–11). Mapping of restriction sites in cpDNA of *Heterogaura* and in eight sections of *Clarkia* demonstrated that *Heterogaura* cpDNA resembled the cpDNA of species in sect. *peripetasma* and not that of the other sections (unpublished data).

MATERIALS AND METHODS

Leaves from 4- to 6-week-old plants of the eight diploid species of sect. *peripetasma*, *H. heterandra*, *Clarkia xantiana* (sect. *phaseostoma*), and *Clarkia amoena* (sect. *rhodanthos*) were frozen in liquid nitrogen and powdered with mortar and pestle. Total DNA was extracted by a described protocol (12) modified to include a 5% polyvinyl-polypyrrolidone grinding buffer and a buffer/powder weight ratio (ml/g) of 40:1. The DNAs were digested with 29 restriction endonucleases that recognize 6-base nucleotide sequences and were subjected to electrophoresis in 0.6%, 0.8%, and 1.0% agarose/Tris/EDTA/acetate gels. After denaturation and neutralization, the fragments were blotted (13) onto BioDyne membranes in a bidirectional fashion. Membrane filters were individually and sequentially probed with each of 12 *Pst I* and two *Sal I* clones, representing the entire cpDNA genome of *Petunia* (kindly provided by J. Palmer and E. Clark). Also used were clones from *Lactuca* and *Phaseolus* (kindly provided by J. Palmer and R. Jansen), which represent distinct portions of the small single-copy region between the inverted repeats of the cpDNA molecule. Restriction sites for 7 enzymes (*Pst I, Kpn I, Sal I, Sma I, Pvu I, Nru I, Xho I*) were completely mapped. Restriction site losses and gains for the additional 22 enzymes were detected from fragment patterns.

For restriction site mutations, the computer program "Phylogenetic Analysis Using Parsimony" (PAUP, version 2.3), developed by D. Swofford (Illinois Natural History Survey, Urbana) was used to find the shortest phylogenetic trees; i.e., those requiring the fewest convergent or back-mutations. The trees were rooted by using *C. xantiana* and *C. amoena* as out-groups. *C. xantiana* possesses the phosho-

The publication costs of this article were defrayed in part by page charge payment. This article must therefore be hereby marked "advertisement" in accordance with 18 U.S.C. §1734 solely to indicate this fact.

5554

Abbreviation: cpDNA, chloroplast DNA.

*Present address: Department of Botany, University of Wisconsin, Madison, WI 53706.

†To whom reprint requests should be addressed.
glucose isomerase gene duplication and is assigned to a related section. *C. amoena* lacks the duplication and belongs to a section thought to represent the ancestral stock of the genus (14). This PAUP program is based on Wagner parsimony (15) and therefore will treat convergent site gains equally with convergent site losses. Dollo parsimony (16), which discriminates against convergent site gains and favors convergent site losses, was also used to construct trees. The computer program ‘Phylogeny Inference Package’ (PHYLIP, version 2.7) with the Dollo parsimony option, developed by J. Felsenstein, was used. The distance algorithm of Fitch and Margoliash (17) was used to construct unrooted networks based on overall measures of cpDNA sequence divergence.

**RESULTS AND DISCUSSION**

**Restriction Site Variation.** Approximately 605 restriction sites (6-base sequences) were surveyed in each of the cpDNAs, for a total sample of 2.1% of the chloroplast genome (172 kbp in *Clarkia biloba*). The *Clarkia* and *Heterogaura* cpDNAs proved collinear to the *Petunia* cpDNA. They did not share a large 45-kilobase-pair inversion characteristic of several species of *Oenothera* (18), also in Onagraceae. A total of 119 mutations were identified among the eight species of sect. *peripetasma* and *H. heterandra*. The highest incidence of change occurred in portions of the large single-copy region near the inverted repeat, a region where a high frequency of mutations has been noted (7, 9). A single mutation was identified within the inverted repeat, confirming its high degree of conservation (6, 9).

**Genetic Distance Measures.** The proportion of substitutions per nucleotide position, *p* (19), estimated from the proportion of site differences between any pair of the eight species in *Clarkia* sect. *peripetasma* and *H. heterandra*, ranged from 0.0017 to 0.0156, with an average of 0.0107 (Table 1). This value is 3–3.5 times higher than that of other cpDNAs within a genus in Brassicaceae, Gentianaceae, and Poaceae (9, 11, 20). The sequence divergence between *H. heterandra* and each of the eight *Clarkia* species ranged from 0.0048 (with *C. dudleyana*) to 0.0147 (with *C. rostrata*), the average being 0.0103 (Table 1). The value for the former pair is the third lowest of all 36 pairwise comparisons, demonstrating that *H. heterandra* and *C. dudleyana* are more closely related to each other than are nearly all pairs of *Clarkia* species tested. Lower values were found only in the progenitor/derivative species pair *C. biloba/C. lingulata* (21, 22) and the morphologically similar *C. cylindrica* and *C. lewissii* (23).

**Phylogenetic Inferences from cpDNA Data.** Fifty-five of the 119 restriction-site mutations were shared by at least two but not all species examined and, consequently, were phylogenetically informative in parsimony analysis. The PAUP program found a single most (Wagner) parsimonious tree that required only six additional mutations to account for all the observed data (Fig. 1). These included two parallel site gains, three convergent or parallel site losses, and one gain/loss (24, 25), for a rate of parallelisms or convergences of 4.8%. This value is nearly 2 times the average rate obtained in *Brassica* (8, 9), *Lycopersicon* (6), and *Pisum* (10). If cpDNA evolution is at all uniform in rate, this suggests a greater age for the *Clarkia/Heterogaura* lineage.

This tree of 125 steps had a consistency index value (26) of 0.95 on a scale from 0 to 1.0, where 1.0 means perfect consistency. It placed *H. heterandra* firmly alongside *C. dudleyana* and not at the base of the phylogenetic tree of *Clarkia* where a distinct genus would be expected. The two species shared nine derived mutations, of which four were unique to them. In all, 16% of the phylogenetically useful mutations were shared by the two species. Thus, they have a more recent common ancestor than either has with any other species tested. Although the species differed by 17 mutations, this difference between them was small relative to the total differences among all species within *Clarkia* sect. *peripetasma* plus *H. heterandra*. The derivation of *Heterogaura* from a common ancestor with *C. dudleyana* was also indicated by the next three most likely (but longer) trees. These trees required 126, 127, and 128 mutations, respectively. In each of them, the close relationship of *Heterogaura* and *C. dudleyana* was maintained. The trees differed only in the relative placement of larger branches or in the placement of *Clarkia modesta* (data not shown).

These Wagner parsimony trees require six or more additional mutations (convergences), two of which are unlikely convergent gains (16, 24, 25). Dollo parsimony, which does not allow for convergent gains, gave a more parsimonious tree exactly congruent in topology with the Wagner tree. This tree, however, required 10 convergent losses or gain/losses (four additional mutations to account for the two convergent gains in the Wagner parsimony tree). The close relationship of *C. dudleyana* and *H. heterandra* is maintained by both the Wagner and Dollo parsimony methods.

Phylogenetic analysis based solely on overall measures of divergence gave exactly concordant results with the analysis based on parsimony. This was demonstrated by utilizing the *p* values in Table 1 in the Fitch and Margoliash algorithm (17). The result was a network of relationships topologically consistent with the phylogeny in Fig. 1 (27). The congruence of results from Wagner parsimony, Dollo parsimony, and distance analyses strengthens our proposed phylogenetic model.

---

**Table 1. Sequence differences among cpDNAs from Clarkia sect. peripetasma and H. heterandra**

<table>
<thead>
<tr>
<th>Species</th>
<th>LEW</th>
<th>CYL</th>
<th>MOD</th>
<th>DUD</th>
<th>LIN</th>
<th>BIL</th>
<th>HET</th>
</tr>
</thead>
<tbody>
<tr>
<td><em>C. epilobioides</em></td>
<td>23</td>
<td>45</td>
<td>40</td>
<td>42</td>
<td>41</td>
<td>49</td>
<td>47</td>
</tr>
<tr>
<td><em>C. rostrata</em></td>
<td>0.65</td>
<td>50</td>
<td>45</td>
<td>47</td>
<td>46</td>
<td>54</td>
<td>52</td>
</tr>
<tr>
<td><em>C. lewissii</em></td>
<td>1.29</td>
<td>1.44</td>
<td>15</td>
<td>41</td>
<td>42</td>
<td>48</td>
<td>48</td>
</tr>
<tr>
<td><em>C. cylindrica</em></td>
<td>1.14</td>
<td>1.29</td>
<td>0.42</td>
<td>36</td>
<td>35</td>
<td>43</td>
<td>43</td>
</tr>
<tr>
<td><em>C. modesta</em></td>
<td>1.20</td>
<td>1.35</td>
<td>1.17</td>
<td>1.02</td>
<td>21</td>
<td>25</td>
<td>25</td>
</tr>
<tr>
<td><em>C. dudleyana</em></td>
<td>1.17</td>
<td>1.32</td>
<td>1.20</td>
<td>0.99</td>
<td>0.59</td>
<td>28</td>
<td>28</td>
</tr>
<tr>
<td><em>C. lingulata</em></td>
<td>1.41</td>
<td>1.56</td>
<td>1.38</td>
<td>1.23</td>
<td>0.70</td>
<td>0.79</td>
<td>6</td>
</tr>
<tr>
<td><em>C. biloba</em></td>
<td>1.35</td>
<td>1.50</td>
<td>1.38</td>
<td>1.23</td>
<td>0.70</td>
<td>0.79</td>
<td>0.17</td>
</tr>
<tr>
<td><em>H. heterandra</em></td>
<td>1.32</td>
<td>1.47</td>
<td>1.29</td>
<td>1.14</td>
<td>0.67</td>
<td>0.48</td>
<td>0.93</td>
</tr>
</tbody>
</table>

The number of restriction site mutations between any two cpDNAs appears in the upper right portion of the matrix. Total number of restriction sites surveyed in each species is *~605*. Values of divergence (19) appear as 100 × *p* in the lower left portion of the matrix.
The problem remains whether the "sister-species" relationship of *C. dudleyana* to *H. heterandra* as seen in the three most parsimonious trees is significantly better in a statistical sense than the placement of *H. heterandra* elsewhere in the tree. Less parsimonious trees that place the *Heterogaura* lineage above or below node "a" (see Fig. 1) are each four mutations longer than the most parsimonious tree. Felsenstein (28) has shown for a clocklike three-species tree that a tree must be 5-10 steps worse before it is significantly worse. Statistical discrimination among these three trees is thus not possible yet. However, placement of the *Heterogaura* lineage below all species within sect. *peripetasma* requires 9 extra steps and can be statistically rejected by Templeton's algorithm (24, 25) as modified by Felsenstein (28). It is clear that although the sister-species relationship of *C. dudleyana* and *H. heterandra* cannot be statistically supported (despite their sharing of a number of derived restriction site mutations), *Heterogaura* can be firmly placed within a specific subsection (a group of four species) of *Clarkia* sect. *peripetasma*.

**Implications for Systematic Biology.** The discovery that the morphologically distinct *Heterogaura* is a *Clarkia*, albeit an unusual one, is noteworthy. But, even more remarkable is that *C. dudleyana* is more closely related to *H. heterandra* than it is to any other *Clarkia* species. The finding confirms the recent speculation that *Heterogaura* might be a derivative of *Clarkia* (4). Morphological evidence, however, could not place *Heterogaura* near any particular *Clarkia* species because their morphological divergence has completely obscured their relationship.

Several important taxonomic implications follow from these results. First, primary reliance on morphological data to model phylogenetic relationships may be misleading no matter how many characters are examined. Discrepancies between morphological and biochemical evidence have also been uncovered in *Brassica* (9), *Lisianthius* (11), and *Tetramolopium* (29). Second, mono- and ditypic genera need not stand in a basal position with respect to related and more speciose genera, but may be derived from within them. Mono- and ditypic genera are not infrequent within Angiosperm families; for example, 5 of the 17 genera in the Onagraceae are of this type (4). Third, the relatively simple and conceptually straightforward restriction enzyme analysis of cpDNA is particularly suitable for testing relationships at the level of the genus and perhaps also the tribe. We conclude that increased utilization of biochemical and molecular evidence will improve the reliability of phylogenetic models in plants but also will pose new dilemmas for classical taxonomic theory.

We are grateful to Rob DeSalle and Bruce Benz for computer assistance. We appreciate comments on the manuscript by three anonymous reviewers. This research was supported by National Science Foundation Grant BSR83-15124.