Recombination in the Chloroplasts of the Florally Diverse Andean Subtribe Iochrominae (Solanaceae)

Erin Collier-Zans

University of Colorado Boulder, collierz@colorado.edu

Follow this and additional works at: https://scholar.colorado.edu/honr_theses

Part of the Computational Biology Commons, Evolution Commons, Genomics Commons, and the Plant Breeding and Genetics Commons

Recommended Citation

https://scholar.colorado.edu/honr_theses/978

This Thesis is brought to you for free and open access by Honors Program at CU Scholar. It has been accepted for inclusion in Undergraduate Honors Theses by an authorized administrator of CU Scholar. For more information, please contact cuscholaradmin@colorado.edu.
Recombination in the chloroplasts of the florally diverse Andean subtribe *Iochrominae* (Solanaceae)

Erin C.E. Collier-Zans
Department of Ecology and Evolutionary Biology
University of Colorado Boulder

Defense Date: October 30, 2015

Advisor: Dr. Nolan Kane

Committee Members:
Dr. Nolan Kane, Department of Ecology and Evolutionary Biology
Dr. Barbara Demmig-Adams, Department of Ecology and Evolutionary Biology
Dr. Robin Dowell, Department of Molecular, Cellular, and Developmental Biology
Abstract

Chloroplast DNA (cpDNA) is a traditional workhorse for reconstructing evolutionary relationships among angiosperms. The frequent use of cpDNA in such analyses is predicated on the apparent simplicity of its inheritance: uniparental through the maternal line, and lacking biparental recombination\(^1\). In hybrid zones, where reproductive barriers between sympatric\(^2\) species may not be fully developed, infrequent leakage of divergent male cytoplasm\(^3\) into hybrid offspring may result in at least transient heteroplasm\(y^4\). Heteroplasmy provides the potential for detectable recombination between maternal and paternal chloroplast genomes to occur. Despite the widespread occurrence of paternal leakage\(^5\) of organelles and heteroplasmy in higher plants, no documented examples of cpDNA recombination in natural angiosperm populations are known. This study presents evidence for the recombination of chloroplast DNA in the Andean subtribe \textit{Iochrominae} (Solanaceae), several species of which are believed to be interspecific or intergeneric hybrids. Recombination was detected by seven distinct methods and verified by approximately unbiased (AU) and Shimodaira-Hasegawa (SH) tests for tree incongruence. The results of this study suggest that a single bifurcating\(^6\) evolutionary history of the plastome cannot be assumed in all lineages. The importance of recombination in angiosperm chloroplasts has broad implications across several fields, ranging from the compromise of phylogenetic reconstruction and the application of molecular clocks to the optimization of agricultural productivity.

\(^1\) Generation of novel genetic material by exchange of DNA between parental genomes
\(^2\) Coexisting in the same geographic area
\(^3\) All material within a cell including its organelles and excluding the nucleus
\(^4\) The presence of two or more organellar genome haplotypes within a cell
\(^5\) In plants, inheritance of pollen-parent cytoplasm by offspring
\(^6\) A phylogenetic tree having two descendants arising from each interior node
Table of Contents

1. Introduction ................................................................................................................................. 1

2. Background ................................................................................................................................. 4
   2.1 Chloroplast Biology .................................................................................................................. 4
      2.1.1 Plastome structure, size, composition, and conservation ................................................. 4
      2.1.2 Chloroplast replication and inheritance........................................................................... 5
   2.2 Recombination in organellar DNA ......................................................................................... 8
   2.3 Review of methods for the detection of recombination ......................................................... 11
   2.4 Characteristics of Iochrominae............................................................................................... 15

3. Materials and Methods ................................................................................................................ 18
   3.1 Alignment and genotyping ...................................................................................................... 18
   3.2 Summary statistics and indirect tests for recombination ....................................................... 21
   3.3 Direct tests for recombination and breakpoint identification .............................................. 22
   3.4 Phylogenetic analysis ............................................................................................................ 23

4. Results ........................................................................................................................................ 25
   4.1 Alignment and genotyping ...................................................................................................... 25
   4.2 Summary statistics and indirect tests for recombination ....................................................... 26
   4.3 Direct tests for recombination and breakpoint identification .............................................. 27
   4.4 Phylogenetic analysis ............................................................................................................ 28

5. Discussion .................................................................................................................................... 28

6. Conclusions and Future Research ............................................................................................... 34

7. Figures and Tables ........................................................................................................................ 35
   Table 1. Taxon information .......................................................................................................... 35
   Figure 1. Significance of pairwise homoplaspy indices across the LSC ..................................... 37
   Figure 2. Results of GARD analysis ............................................................................................ 38
   Table 2. Summary of non-recombinant fragment statistics ....................................................... 39
   Table 3. Results of GENECONV analysis of gene conversion .................................................. 40
   Figure 3. Gene conversion events relative to the Physalis peruvianum LSC ............................... 41
   Figure 4. Approximate Unbiased and SH tests for phylogenetic incongruence ....................... 42

8. Acknowledgements ...................................................................................................................... 43

9. References .................................................................................................................................... 44

10. Supplementary Material ............................................................................................................ 49
   Sup. Figure 1. Phylogeny of Iochrominae based on three nuclear loci ....................................... 49
   Sup. Figure 2. Global median read coverage for 67 taxa ........................................................... 50
   Sup. Figure 3. Physalis peruvianum reference genome and SNP density map for 67 taxa ........ 51
   Sup. Figure 4-21. GARD non-recombinant fragment phylogenies ............................................ 52
1. Introduction

The field of plant molecular systematics relies on sequence data from the nuclear, mitochondrial, and chloroplast genomes for the inference of phylogenetic relationships. Inference based on data from the nucleosome may be subject to a host of potential confounding factors associated with the frequently complex evolutionary histories of nuclear loci, histories that are not shared by the relatively simple genomes of organelles. Mitochondria and chloroplasts are endosymbiotes of prokaryotic origin that retain small, highly reduced genomes independent of the nucleosome. Organelles in most eukaryotes are inherited from a single parent, and in the majority of land plants the mother donates cytoplasmic material to offspring. Both mitochondria and chloroplasts have long been believed to be incapable of recombination under natural conditions, and in this traditional paradigm the mitochondrial and chloroplast genomes are each considered a single linkage group. Homoplasy may be expected to introduce weak signal of phylogenetic conflict, but if homoplasy is accounted for, analysis of any organellar DNA segment should reveal the bifurcating phylogenetic history of the entire organelle genome.

The longstanding view that mitochondrial DNA is an ideal tool for systematics in eukaryotes has recently been cast in doubt. A growing body of evidence supports horizontal gene transfer (hereafter referred to as recombination) in the mitochondria of diverse vertebrates, invertebrates, fungi, and plants, and this development may lead to reconsideration of the phylogenetic utility of organelles, including chloroplasts. There are very few documented cases of recombination among the chloroplast genomes of flowering plants, and only in the chloroplasts of artificially induced somatic crop hybrids or “cybrids” (Medgyesy et al. 1985; Houliston and Olson 2006), though it has been suggested that homologous recombination may be partially responsible for the maintenance of chloroplast genome structure and stability (Marechal
and Brisson 2010). There are only two documented cases of recombination in the chloroplasts of non-agricultural species (Huang et al. 2001; Marshall et al. 2001), and both of these observations were made in gymnosperms.

A history of recombination or horizontal gene transfer\(^7\) within a genome may manifest as incongruence between gene trees. Incongruence may be of two types: hard or soft. Soft incongruence is disagreement between poorly supported trees inferred from sequences having a shared evolutionary past. This form of incongruence is typically due to inappropriate model selection, insufficient data leading to lack of phylogenetic signal, or rampant homoplasy. It can often be mitigated by inclusion of more informative data or selection of a model that accounts for homoplasious traits. Only with sufficient informative data can hard incongruence due to lack of commonality in evolutionary history be detected. If hard incongruence is present in sequence data, true gene trees resolved from different loci will present conflicting relationships. There are no reported instances of hard incongruence in chloroplast DNA (cpDNA).

Early studies in plant systematics utilizing cpDNA were typically based on a single gene. The biotechnological advancements of recent years have facilitated inclusion of additional genes in multi-loci analyses, which often offer greater power to determine relationships amongst taxa. For instance, Graham and Olmstead (2000) based a phylogenetic analysis of basal angiosperm relationships on 17 plastid genes (13.8-Kbp), Olmstead et al. (2008) made steps towards resolving the phylogeny of Solanaceae using two genes sampled from 195 taxa, and Goremykin et al. (2003) made a study from the sequences of 13 fully assembled chloroplast genomes. Only 20 complete chloroplast genomes were publicly available in 2004 (Wolfe and Randle 2004); that number has since jumped to more than 600.

\(^7\) The process of asexual gene transfer between contemporary organisms, most often occurring in prokaryotes.
Because of the afore-mentioned properties of the chloroplast genome, loci from different regions of the plastome are typically concatenated\(^8\) without evaluation for signal of incongruence (e.g. Bremer et al. 2002; Nishiyama et al. 2004). Conflict among different regions of the chloroplast are typically treated as artifacts or attributed to homoplasy\(^9\). Given the assumption that the chloroplast represents a single linkage group evolving in a bifurcating fashion along a single tree topology, poorly resolved phylogenetic trees should be a consequence of either poor model selection or lack of phylogenetically informative characters. Few to no attempts are made in studies utilizing chloroplast DNA to rule out recombination prior to analysis or as a potential contributor to conflict among gene trees, and as such the traditional view that chloroplasts are non-recombinant in naturally occurring populations of angiosperms has gone unchallenged.

This study uses 86-Kbp from the long single copy (LSC) of 67 individuals from the Andean subtribe *Iochrominae* (Solanaceae). The LSCs of 29 of 35 classified species, 4 unpublished species, and 6 interspecific hybrids were analyzed. Chloroplast genotypes were determined by aligning whole-plastome Illumina shotgun reads to a published reference genome (*Physalis peruvianaum*). Sequence polymorphisms were then screened for evidence of recombination in a three-phase hierarchical fashion, first with methods designed to detect recombination’s presence or absence both with and without the assumption of homoplasy, followed by detection of recombination breakpoints, and finally with topology tests for phylogenetic conflict between gene trees generated from non-recombinant sequence fragments. This study concludes that there is significant evidence of recombination in the chloroplasts of the angiosperm subtribe *Iochrominae*.

---

8 To combine strings of characters (e.g. segments of sequenced DNA) end-to-end
9 A character shared by a set of species that is not present in their last common ancestor, typically due to convergent evolution
2. Background

2.1.1 Plastome structure, size, composition, and conservation

Chloroplasts and other proplastid-derived organelles are believed to have arisen from a shared cyanobacterial progenitor (Goksoyr 1967; Grey and Doolittle 1982) through a process of primary endosymbiosis estimated to have occurred approximately 1 billion years ago (Reyes-Prieto et al. 2007; Gould et al. 2008; Archibald 2009). The precise lineage of chloroplasts remains uncertain, but their closest extant relatives are believed to be N$_2$ fixing unicellular cyanobacteria belonging to the order Chroococcales (Falcon et al. 2010). Adaptation of ancestral chloroplasts to endosymbiosis involved a process of extreme genome reduction and gene loss. Modern plastid genomes typically range in size from 135-160-Kbp, while modern cyanobacterial genomes vary in length between 1.7 and >9-Mbp (Dufresne et al. 2003; Timmis et al. 2004).

Chloroplast genome structure, sequence, gene content, and gene order is highly conserved within land plants. With few exceptions, chloroplast genomes are comprised of a large inverted repeat approximately 25-Kbp in length segregating remaining sequence into long and short single copy (LSC, SSC) regions (Palmer 1985). Changes in gene order are rare, and typically arise through relatively small-scale sequence inversions or by gene loss (Marechal and Brisson 2010). Plastids in most lineages including land plants retain a very small complement of genes, the precise number of which varies amongst plastid types, between plant groups, and according to life history. The 60-200 genes encoded by typical angiosperm chloroplasts belong to primarily photosynthetic lineages. Ubiquitous genes are typically those involved in photosystem assembly (psa, psb, and rbc gene families), DNA transcription and translation (rpl, rpo, rps), fatty acid metabolism (acc, fab), ATP production and energy metabolism (atp, ndh, pet), and metabolic cofactor synthesis (acc, chl, ccs, acs). The average angiosperm plastome
encodes 113 products in total, including 4 rRNAs, 30 tRNAs, and around 80 proteins (Shinozaki et al. 1986; Olmstead and Palmer 1994).

The vast majority of the chloroplast proteome is encoded by the nucleosome. Between 2100 and 3600 proteins are estimated to be targeted to chloroplasts (Abdallah 2000). The function of the majority of plastid-targeted genes remains unclear, but many are known to be essential to organelle function (Wolfe et al. 1994; Allen and Forsberg 2001) and are believed to have been transferred to the nucleosome subsequent to endosymbiosis (Martin et al. 1998; Martin et al. 2002; Timmis et al. 2004). Estimates for the total number of genes transferred from plastid ancestors to host genomes remain elusive, but studies in the model organism Arabidopsis thaliana suggest that somewhere between 800 and 2000 genes, and as much as 18% of the nuclear gene compliment, are of cyanobacterial origin (Abdallah et al. 2000; Arabidopsis Genome Initiative 2000; Martin et al. 2002). There are a number of potential benefits of large-scale gene loss and transfer, the most probable of which are reduction of genetic load through eukaryotic recombination and the relatively high rate of mutation experienced by nuclear DNA, which may be between 3 and 10 times higher than rates in the chloroplast and mitochondria, respectively (Drouin et al. 2008).

2.1.2 Chloroplast replication and inheritance

All plastids (chloroplasts, chromoplasts, leucoplasts, amyloplasts, and others) are derived from nongreen proplastids found in cells at the apical meristem of developing tissue (Lopez-Juez and Pyke 2005). Although the mechanism organizing synchronized replication of eukaryotic cells and organelles is poorly understood (Miyagishima 2011), it is known that proplastids, typically numbering between 10 and 20 per cell (Juniper and Clowes 1965), replicate to keep
pace with cell division. As cells expand without division during leaf development, differentiated plastids continue to replicate non-synchronously until the number of chloroplasts per cell reaches between 100 and 200 (Possingham and Lawrence 1983). Each chloroplast contains multiple copies of the plastome, the conformation of which may be either circular\(^\text{10}\) or branching linear\(^\text{11}\) (Bendich 2004).

Plastids, including proplastids and chloroplasts, usually divide by a process of binary fission involving the simultaneous constriction of the inner and outer envelopes at a division site. Electron-dense ring structures termed plastid-dividing (PD) rings have been found at these division sites (Birky 2001). All known components of the division complex in angiosperms are encoded by the nuclear genome (Kuroiwa et al. 2008; Yang et al. 2008; Maple and Moller 2010; Miyagishima and Kabeya 2010). As in cyanobacteria, the division complex proteins FtsZ (filamenting temperature-sensitive mutant Z) and ARC6 (accumulation and replication of chloroplast 6) assemble into a ring positioned on the stromal surface of the organelle’s midpoint by machinery related to the cyanobacterial Min system. The mechanism responsible for the constriction of plastid-dividing rings during binary fission is currently unknown.

During plant reproduction, plastids are transferred to the zygote by the maternal or paternal parent in the majority of angiosperm and gymnosperm lineages, respectively, either in conjunction with or independent of the mitochondrion. Uniparental inheritance of the cytoplasm is commonly thought to have arisen as a nuclear trait counteracting the negative effects of selfish\(^\text{12}\) organellar genomes or cytoplasmic parasites by reducing the competitive advantage of selfish elements and by reducing within-host variation (Hastings 1992; Rispe and Moran 2000).

\(^{10}\) DNA that forms a closed loop with no free ends
\(^{11}\) DNA that does not form a closed loop, and has two or more free ends
\(^{12}\) Organellar traits promoting their own inheritance, but not necessarily the survival of the organism
Only a handful of groups practice uniform biparental inheritance, but up to one third of angiosperm genera studied by Smith (1989) may infrequently inherit plastids from both parents. Destruction or exclusion of the male cytoplasm during pollen and sperm development or immediately prior to syngamy\textsuperscript{13} is the norm in lineages practicing exclusive or near-exclusive maternal inheritance.

Pollen is generated in a multistage-process beginning with the meiosis of diploid microsporocytes in young anther tissue. The four resultant haploid microspores, each containing cytoplasmic organelles, mitotically divide to produce one vegetative and one smaller generative cell each. Plastids may be excluded from the generative cell during mitotic division of the microspore via polarization (Hagemann and Schroder 1989) mediated by microtubules (Van Went 1984; Tanaka 1991), actin filaments (Schroder et al. 1988; Pierson and Cresti 1992), or biochemical gradients (Schroder 1985). If plastids are not excluded during primary mitosis they may be inherited by sperm cells during the mitotic division of generative cells (Schroder and Oldenburg 1990). Plastids that escape polarization may be subsequently excluded from generative or sperm cells during later mitotic divisions. Male cytoplasm may also be excluded during gametic fusion by enucleation, in which case the male cytoplasm remains appressed to the exterior surface of the egg at the point at which the sperm’s nucleus entered (Birky 2001).

Even if male cytoplasmic DNA escapes multiple stages of suppression to be inherited by the zygote, it may be nonfunctional. Studies in albino and chimeric\textsuperscript{14} plantlets revealed that zygotes inheriting male cytoplasm often possess severely deleterious photosystem mutations (Day and Ellis 1984; Day and Ellis 1985; Dunford and Walden 1991; Harada et al. 1991) leading to their mutant phenotypes. This pattern suggests that modification of cpDNA in the form of

\textsuperscript{13} Fusion of gamete nuclei during eukaryotic reproduction

\textsuperscript{14} Having two or more populations of genetically distinct cells
large-scale deletion may take place during pollen maturation. These modifications render chloroplasts from the male cytoplasm ultimately uninheritable. And yet, despite the multi-fold suppression of male cytoplasm in angiosperms, there remain recorded instances of biparental inheritance resulting in the paternal plastid being present in a small percentage of progeny (Medgyesy et al. 1986; Schmitz and Kowallik 1986; Cornu and Dulieu 1988; Horlow et al. 1990; Sewell et al. 1993). This phenomenon is commonly referred to as paternal “leakage”.

2.2 Recombination in organellar DNA

Genetic recombination and nucleotide substitution are two processes by which the variation upon which natural selection acts is generated. Sexual recombination in eukaryotes and horizontal gene transfer in prokaryotes involve the transfer of genetic material between genomes, though genetic material may also be transferred within genomes or chromosomes. Recombination is considered homologous when sequences replace others having a shared evolutionary history, and recombination is said to be nonhomologous when it involves the joining of unrelated sequences. Homologous recombination plays a central role in the cellular and organellar repair of broken or damaged DNA molecules, and is pivotal to the replication and repair of chloroplast DNA (Marechal and Brisson 2010).

When recombination is homologous, variation is introduced to genomes in the form of novel alleles. Populations sharing alleles typically have greater variation than do non-recombining populations, and the former are thereby often better equipped to respond to natural selection than the latter. Without recombination, deleterious mutations arising alongside beneficial mutations gradually accumulate, even in the fittest of genomes, in a process called
Muller’s ratchet (Vos 2009). As the genetic load\(^\text{15}\) of a population increases, overall fitness\(^\text{16}\) decreases. Therefore, in a fitness landscape\(^\text{17}\) with some fitness optimum, Muller’s ratchet eventually results in the population sliding down the fitness peak. Populations with small effective population sizes, including endosymbionts (Herbeck et al. 2003), may experience repeated genetic bottlenecks and run the risk at every constriction of losing their fittest genotypes. Homologous recombination counteracts the process of fitness slippage\(^\text{18}\) by reconstruction high-fitness, deleterious-mutation-free individuals, but recombination’s benefits do not come without a cost. The genetic “load” of recombination is the loss of overall population fitness as a result of the breakup of beneficial allele combinations (Otto and Lenormand 2002). Beneficial combinations existing in the present have survived many rounds of natural selection and may, on average, be fitter than random combinations produced by recombination. This fitness differential is most extreme when genotypes are near the peak of the fitness landscape (Vos 2009).

Mitochondria and chloroplasts have traditionally been believed to be both nonrecombinant and subject to strong purifying selection, indicating that organellar genotypes not occupying the peak of their fitness landscape may swiftly be removed from the population. If organelles are also subject to Muller’s ratchet, the fittest class of endosymbionts would run the risk with each new generation of being lost to accumulation of deleterious mutations or failure to be transmitted into gametes. In the absence of recombination, the process of Muller’s ratchet in organelles could only be avoided if back or compensatory mutations arose at sufficiently high rates to balance or offset it. There is indication that the rate of deleterious mutation accumulation

\(^{15}\) The difference between the average observed genotype fitness and the theoretical fitness optimum

\(^{16}\) Individual reproductive success

\(^{17}\) A tool for visualizing genotypic fitness, where troughs are occupied by individuals with poor fitness and peaks by individuals with high fitness

\(^{18}\) Decrease in average population fitness corresponding to increase in genetic load
in organellar genomes may be sufficiently low that compensatory mutation could, in fact, outweigh it (Lynch and Blanchard 1998), but infrequent recombination may also help to reduce genetic load (Barr et al. 2005). Relatively high sequence diversity and rates of recombination are required for the detection of recombination by most modern methodologies (Maynard Smith 1999; Posada and Crandall 2001; Posada 2002; Posada et al. 2002; Wiuf 2001), and organellar recombination, if it occurs, would most often be between highly similar sequences contained within a single organelle or cell. This suggests that true rates of organellar recombination cannot be accurately estimated by current methodologies given its probable undetectability.

Detection of recombination in any given sequence alignment requires that, at minimum, recombinant fragments differ by at least two phylogenetically informative sites (Martin et al. 2011). Given that this prerequisite is met, the strength of signal will increase as: 1) the number of variable sites differentiating parental sequences increases, 2) the relative proportions of variable sites contributed by each parent becomes more even, 3) phylogenetic distance between parental sequence increases, and 4) sampling becomes more exhaustive (Martin et al. 2011). Ideally, data subjected to analysis should contain both parental and recombinant sequences, but if parental taxa are absent other sequences bearing greater resemblance to one of the parental sequences than the other must be present (Martin et al. 2011). As parental sequences become more distant (e.g. as in the case of intespecific or intergeneric hybrids), the signal for recombination increases. As mutations accumulate in both the recombinant and parental lineages over time, the signal for recombination will erode. Given that sequences on either side of a recombination breakpoint must have at least two phylogenetically informative sites differentiating them, at least transient heteroplasmy must occur for the recombination of organellar genomes to be detectable (Birky

---

19 Single nucleotide substitutions or mutations shared by two or more individual sequences
Heteroplasm in organellar genomes is not uncommon (Wagner et al. 1991; Barr et al. 2005; Azhagiri and Maliga 2007; Ellis et al. 2008; Pearl et al. 2009; Bentley et al. 2010; Nunes et al. 2013), and may arise either by point mutation or by paternal leakage resulting in the presence of both maternal and paternal cytoplasm in the cells of offspring.

**2.3 Review of methods for the detection of recombination**

Most methods designed for the detection of recombination do so by determining whether either phylogenetic relationships or relative degrees of similarity between sequences in an alignment vary in a manner consistent with recombination. Methods can then be broken down into two additional categories: parametric\(^ {20} \) or nonparametric\(^ {21} \). One of the earliest non-parametric methods for detecting recombination is the four gamete test of Hudson and Kaplan (1985). The four gamete test searches for incompatible pairs of variable sites within a parsimony framework. Incompatible sites in this context are those displaying patterns of variation supporting phylogenetic incongruence. Given two sites, each with two alleles, we find four possible site combinations: 00, 01, 10, and 11. A pair of sites is only said to be incompatible when all four combinations (“gametic types”) exist; three or fewer combinations can be explained by mutation alone, without invoking recombination. As with most tests using the conceptual framework of maximum parsimony, the four gamete test is susceptible to error in the form of false positives where homoplasy due to parallel, convergent, or reverse-substitution mutation is present (Martin et al. 2011).

---

\(^ {20} \) Having to do with the estimation of population parameters (e.g. nucleotide substitution rates)
\(^ {21} \) Distribution-free methods
The MaxChi2 program of Maynard Smith (1992) introduced the Maximum $X^2$ test to recombination detection. Subsequent implementations of Max $X^2$ (e.g. PhiPack) differ little in theory from MaxChi2’s original execution. For every pair of sequences, the maximum chi-squared statistic is calculated on a sliding window of variable sites that is advanced one nucleotide at a time. The width of this window is arbitrarily set to the total number of variable sites in the alignment divided by 1.5. The p-value for the null hypothesis of no recombination is estimated as the proportion of times the maximum chi-square was smaller than the maximum chi-square of 1,000 (or a user-defined) number of permuted alignment replicates, obtained by randomizing relative site positions in a column-wise fashion.

The Neighbor Similarity Score or NSS (Jakobsen and Easteal 1996) utilizes compatibility matrices of adjacent informative sites to test for signal of recombination. Phylogenetically informative site pairs are determined to be compatible if the minimum number of possible changes $c$ is one less than the number of distinct nucleotides $n$ at each site, and incompatible if $c > n - 1$. NSS values may be higher than expected by chance due to recurrent mutation, gene conversion, or recombination (White and Gemmell 2009). Significance of the NSS statistic is determined in a similar fashion to Max $X^2$, by permuting the relative order of informative sites 1,000 or more times and determining the fraction of random NSS scores greater than or equal to the observed score.

The Pairwise Homoplasy Index or PHI test of Bruen et al. (2006) is a relatively recent development in the field of indirect tests for recombination, and has quickly gained widespread acceptance due to its robustness against homoplasy and substitution rate heterogeneity (Bruen et al. 2006; White et al. 2013). PHI measures the mean refined incompatibility between sites within a sliding window with user-defined width and step size. In the absence of recombination, the
refined incompatibility score between any two sites reflects the minimum number of homoplasies that have occurred in the evolutionary history of the site pair. In the opposite case, i.e. the absence of homoplasy and the presence of recombination, the RIS represents the minimum number of recombination events in the data set assuming recombination presents itself as a phylogenetic unrooted subtree-prune and regraft operation. The statistical significance of PHI is achieved, again, by randomly permuting relative site positions to simulate data lacking recombination, and calculating the proportion of times permuted PHI scores met or exceeded the observed score.

The four gamete test and the three indirect methods for detection of recombination described above have been utilized in this study. Indirect methods have the advantage of providing a single estimate of significance; relatively weak signals of recombination disbursed across the span of an alignment may accumulate such that indirect methods detect recombination events that would be missed by more direct methods. The downside of most indirect methods is their lack of model for nucleotide substitution, making them more susceptible to false positives than parameterized direct methods where homoplasy is abundant.

Two direct methods for breakpoint detection are utilized in this study: SBP and GARD. The Rapid Screening for Recombination Using a Single Break Point (SBP) analysis is a powerful method for the detection of a single breakpoint within a data set (Kosakovsky Pont et al. 2006). SBP accepts a user-defined reversible model of nucleotide substitution, and assumes that in the absence of recombination the evolutionary history of a data set should be uniform across the length of the alignment. At the outset, a neighbor joining tree is fitted to the complete data set and an Akaike score derived in an ML (Felsenstein 1981) framework for rate parameters and branch lengths. Informative sites are subsequently divided in a stepwise fashion into V - 1
continuous blocks, where $V$ coincides with the number of variable sites and each variable site is considered a putative breakpoint. An NJ tree is computed independently on either side of each putative breakpoint and an Akaike score is calculated for the model fitting branch lengths to each partition. If a partition Akaike score is found to be less than the Akaike score computed for the full data set, the algorithm deduces that at least one sequence in the alignment is recombinant. Relative support for every putative breakpoint is then gauged by computing the $\Delta AIC_c$ score for every adjacent breakpoint until the most optimal breakpoint is found.

The Genetic Algorithm for Recombination Detection or GARD (Kosakovsky Pond et al. 2006) follows a similar procedure for the identification of multiple breakpoints, minimizing computational costs by utilizing an aggressive population-based hill-climbing algorithm to search the space of candidate breakpoint models for the optimum. Parameter space for this optimization problem has two components: a discrete allocation of sites in the alignment to $B$ breakpoints and a vector of real parameters corresponding to branch lengths. The hill-climbing algorithm is used to hasten the costly search of discrete parameter space, and ML is used to estimate all other model parameters. The fitness of every model is measured by its $AIC_c$ score. NJ trees are constructed for every fragment segregated by $B$ breakpoints, and $B$ is incrementally increased by 1 until the $AIC_c$ score of the best model remains stationary for 100 generations, at which point the analysis is stopped and the optimal breakpoint model declared. GARD does not explicitly require that tree topologies differ among sequence partitions, so phylogenetic incongruence tests were used both by Kosakovsky Pond et al. (2006) and this study to verify topological disagreement between sequence fragments partitioned at $B$ optimal breakpoints.

Direct methods for breakpoint detection do not themselves provide any information about which sequences in an alignment have donated and received genetic material during homologous
recombination. In order to detect individual recombination events and their participants, the geneconv algorithm of Sawyer (1989) as implemented in GENECONV (Sawyer 1999) was used to search for potential gene conversion events. “Gene conversion” is an umbrella term for processes that cause DNA segments from one molecule to be homologously copied into another molecule. Conversion of short DNA fragments is a powerful force in evolution (Gyllensten et al. 1991; Hilliker et al. 1991), and is vital in many systems (including organelles) for the repair of broken or damaged DNA (Morton and Clegg 1993; Khaklova and Bock 2006).

For every pair of individuals in a data set, GENECONV locates the longest identical or nearly identical aligned DNA segments. Similarity is rewarded by penalizing mismatches and assigning identical bases a score of +1. The bounds of candidate conversion fragments are defined either by the termination of the alignment or by discordant sites, or both. Discordant sites are sites where the pair of aligned individuals do not share the same allele. Both local and global p-values are computed for conversion events by sequence permutation. Local or “pairwise” p-values are the proportion of permuted alignments for which the maximum fragment score for that pair of sequences is greater than or equal to the observed fragment score. Global permuted p-values are more conservative; these are the proportion of all possible fragments in all sequence pairs greater than the observed fragment.

2.4 Characteristics of Iochrominae

Iochrominae is a subtribe of Solanaceae distributed predominantly along the western slope of the Andes from Columbia to Argentina, with the majority of their diversity found in the Amotape-Huanacamba zone (see Supplementary Figure 1) at the border of Ecuador and Peru (Smith and Baum 2006). Most species within this zone occur in sympatric communities of two to
four species, while species outside this zone are typically found in allopatry. Most *Iochrominae* grow in cloud forest gaps and disturbed areas (including trails, dry stream beds, and field edges) between 2200 and 2900 meters in elevation (Smith et al. 2008). Some species are widespread and reasonably cosmopolitan in habitat preference with ranges up to 50,000 km², while other species are narrow endemics with ranges as small as 40 km². *Iochrominae* are distinguished by their woody shrub or tree-like growth habits, lack of chemical armament, and showy, tubular flowers (Smith and Baum 2006; Smith and Baum 2011). *Iochrominae*’s diverse species exhibit all major flower colors and corolla forms found in the wider Solanaceae, with described species having red, orange, yellow, green, blue, purple, and white flowers. Corollas are rotate or tubular and vary up to eight-fold in length between species, with some taxa having tube lengths less than 1 cm and others upwards of 6 cm. Chromosome counts by Hunziker (2001) suggest that all *Iochrominae* are diploid with n = 12.

Together with *Physalinae* and *Withaninae*, *Iochrominae* sensu Olmstead et al. (1999) form the large clade *Physaleae*, which is the sister tribe of *Capsiceae*. The monophyly of *Iochrominae* has been established by morphological phylogenetic analysis (Sawyer 2005), and flower morphology was used historically to divide the then-34 recognized *Iochrominae* species into six traditional genera. These genera are *Acnistus* (Hunziker 2001), *Dunalia* (Hunziker 1960), *Eriolarynx* (Hunziker 2000), *Iochroma* (Shaw 1998), *Saracha* (Alvarez 1996), and *Vassobia* (Hunziker 2001). Molecular phylogenetic analysis using three nuclear loci (ITS, LEAFY intron 2, GBSSI exons 2-9) indicate that these genera are largely unnatural, non-monophyletic groupings, suggesting that many floral traits in *Iochrominae* are homoplasious and have likely emerged by convergent evolution (Smith and Baum 2006). An alternative system of

---

22 The petal whorl
classification of *Iochrominae* has since been suggested, comprised of 9 monophyletic clades: D, E, S, and V for the traditional genera *Dunalia, Eriolarynx, Saracha,* and *Vassobia,* with the addition of A, C, L, F, and U subdivisions of *Iochroma* (Smith and Baum 2006).

Several species of *Iochrominae* have floral morphologies that follow typical pollination syndromes, e.g. red, scentless, tubular flowers typifying hummingbird specialization (*I. fuchsioides*), and white, fragrant flowers characteristic of many species pollinated by nocturnal moths (*I. ellipticum*). Most *Iochrominae,* however, do not possess floral traits placing them definitely in any one syndrome. Smith and Baum (2008) found that the majority of *Iochroma* species are diurnally hummingbird pollinated, with a few species practicing nocturnal or diurnal insect-exclusive pollination and several others utilizing a mix of insect and bird pollinators. Plants producing large nectar rewards are visited almost exclusively by large, territorial hummingbird species, while smaller birds most commonly visit relatively low-reward plants. Territoriality enforces a high degree of plant fidelity on the part of individual hummingbirds, which likely serves as the primary barrier against interspecific pollination in *Iochroma* and *Iochrominae* occurring sympatrically.

Despite this primary reproductive barrier, species of interspecific hybrid origin have been identified through phylogenetic analysis. Smith and Baum (2006) recognized three putative hybrids (*I. ayabacense,* *I. “sagasteguii”,* and *I. stenanthum*) based upon their intermediate phenotypes and conflicting position in nuclear gene trees. In each case, putative parental taxa were not found to be each other’s closest relatives, suggesting wide crossability within the subtribe. These and a small number of other taxa believed to be of hybrid origin have been observed both in collections and in natural communities where multiple species of *Iochrominae* co-occur (S. D. Smith, personal communication, October 2014).
A study on interspecific crossability (Smith and Baum 2007) using seven taxa found that only two of 21 species pairs failed to produce seed in both directions. All remaining pairs produced viable seed in at least one direction. Many interspecific crosses were comparably successful in fruit set, seed set, and germination to intraspecific crosses, though most interspecific crosses resulted in fewer seeds per fruit and several produced fruits containing either no seeds or seeds lacking embryos. This suggests the presence of post-pollination, pre-seed barriers to reproduction between some but not all species pairs. One F1 (*I. gesnerioides* x *I. cyaneum*) was grown to maturity, and these hybrids were able to successfully produce fruit and germinable seed when crossed to either parent, though they displayed low pollen viability. This indicates that some F1s are viable and could intercross with parental populations, though the extent of this viability is uncertain and hybrid fitness in situ remains unclear. The potential for interspecific or even intergeneric hybridization and the paternal leakage of divergent cytoplasmic types presents a unique opportunity for detectable chloroplast recombination to occur.

3. Materials and Methods

3.1 Alignment and genotyping

All sequencing was done by Daniel Gates at the University of Nebraska-Lincoln on an Illumina HiSeq platform. Taxa included 67 individuals belonging to 29 species from the six traditionally recognized genera of *Iochrominae*, four intergeneric hybrid individuals produced by controlled cross of *Iochroma cyaneum* and *Acnistus arborescens*, and five putative natural hybrids (S. D. Smith, personal communication, October 2014) and their parental taxa (Table 1). Both non-capture and capture data enriched for 246 nuclear loci were available, and these data were concatenated to maximize potential coverage of the chloroplast.
Each set of read files was trimmed using Trimmomatic v0.32 (Bolder et al. 2014) with parameters [ILLUMINA_CLIP:illumina.fa:2:20:10 LEADING:20 TRAILING:20 MINLEN:90] to remove sequence adapters and low quality reads and bases. Reads were then aligned to the *Physalis peruvianum* reference (NCBI RefSeq NC_026570.1) using the Burrows-Wheeler Aligner (BWA) v0.7.5a (Li and Durbin 2009) with relaxed parameters, as follows: minimum seed length (-k = 13), mismatch penalty (-B = 2), gap open penalty (-O = 2), and clipping penalty (-L = 3). SNPs were called using FreeBayes v0.9.18 (Garrison and Marth 2012). The ‘--ploidy’ option was used to set the expected number of chromosomes to 1 (haploid). SNP calling required a minimum fraction of 0.8 observations within an individual supporting the alternate allele. Non-SNP polymorphisms were excluded at this stage. FreeBayes by default excludes from consideration reads with mapping quality less than 1.

A proprietary pipeline (Keepers K, Collier-Zans E, Tittes S, unpublished) was used to control for patchy coverage. A list of taxon names was provided to a wrapper script, and this wrapper then carried out the following procedure independently for every individual in the data set: 1) Following read trimming and alignment, SAMtools depth (Li et al. 2009) was used to report the depth of coverage at every site, monomorphic and polymorphic, in the alignment. 2) An AWK script was used to parse the depth file and calculate from it the “global” (alignment-wide) median read depth (see Supplementary Figure 2 for summary of these data). 3) Using the global median statistic and user-defined cutoffs for the minimum multiple, maximum multiple, and absolute floor of coverage, sites in the depth file not meeting requirements for coverage were recorded as dashes (-). 4) Following compilation of dash files, vcf files output by FreeBayes were filtered such that only SNPs meeting a user-defined minimum quality score were retained.

---

23 Site-to-site variance in the number of times a nucleotide was read during the sequencing process
5) During the final stage of the pipeline, the vcf table and dash file were converted into a fasta file. This step took the reference fasta as input, and from it generated an ecotype fasta by i) inserting appropriate alternate alleles at sites indicated by the filtered vcf table, and ii) inserting dashes at every site failing to meet coverage requirements as indicated by the output of step 3. All ecotype fastas were then combined, resulting in a high-quality data set largely free of alignment errors.

For the purposes of this study, minimum and maximum multiples of global median read depth per individual of 1/5 and 6 were required. The pipeline made no distinction between high and low coverage individuals in the data set (e.g. 468 vs. 15 median reads per site), so a compromise had to be made between maximizing available data in low-coverage individuals and minimizing the potential for erroneous SNP calls in a small number of regions with extremely high coverage attributable to sequencing error. Because global median read coverage across the data set was heavily skewed towards the right (Supplementary Figure 2), minimum and maximum multiple parameters were relaxed to maximize data retention. An absolute floor of coverage of 5 reads was set to exclude low coverage sites escaping the minimum multiple of 1/5. Only SNPs meeting or exceeding a minimum quality score of 300 were retained.

Since the inverted repeat is known to be sequentially more conserved than either single copy region (Wolfe et al. 1989; Goremykin et al. 1996), it was assumed that the IR would not have accumulated mutations sufficient to be phylogenetically informative at the intrageneric level (Supplementary Figure 4-21). The IR was clipped from the alignment and excluded from later analyses, as was the short single copy. The boundaries of the long single copy (LSC) were estimated using NCBI BLAST (Altschul et al. 1997; http://www.ncbi.nlm.nih.gov/) in a self-comparison of *P. peruvianum*. 500 bases were clipped from both the beginning and end of the
LSC to account for potential motility of the IR (Aldrich et al. 1998; Goulding et al. 1996; Chung et al. 2006). 86-Kbp of the long single copy was retained. Alignment errors were manually corrected in Mesquite v3.04 (Maddison and Maddison 2015).

A BLAST comparison of 246 nuclear loci to the *P. peruvianum* plastome revealed that 94 of 246 enriched genes shared at least some homology with the chloroplast LSC. To assess potential contamination by nuclear reads, 10 individual samples were selected at random and aligned to a combined reference comprised of the *P. peruvianum* plastid and 94 nuclear homologs using BWA mem with alignment parameters as described above. To count reads mapping non-uniquely to the chloroplast, reads mapping at least once to *P. peruvianum* with a minimum mapping score of 1 were segregated. Grep was used to search SAM alignments for the headers of chloroplast reads mapping to one or more nuclear loci. The number of reads per individual mapping both to the plastid and one or more nuclear loci was small (µ = 8.6, µ̄ = 5.99). With so few reads mapping non-uniquely to the chloroplast reference, alteration of the alignment pipeline was deemed to be unwarranted.

3.2 Summary statistics and indirect tests for recombination

Per-site nucleotide diversity (π), average number of nucleotide differences (k), and scaled mutation rate (θ) were calculated in DnaSP v5.10.1 (Rozas and Rozas 1995; Librado and Rozas 2009) using equations from Nei (1987), Tajima (1983), and Watterson (1975) and Nei (1987), respectively. DnaSP was also used to estimate the recombination parameter (R) and the minimum number of recombination events (RM) according to Hudson and Kaplan (1985) and Hudson (1987). RM is obtained from the results of a four-gamete test of every variable site pair,
and R between adjacent sites is estimated from RM via simulation using the following expression:

\[ R_{adj} = \frac{4Nr}{D} \]

where \( N \) is the effective population size, \( r \) is the recombination rate per generation between the most distant sites, and \( D \) is the average nucleotide distance in base pairs discounting alignment gaps.

The data were also tested for the presence of recombination using the Pairwise Homoplasy Index or PHI (\( \Phi \)) statistic, Max X\(^2\), and NSS. These tests were performed using the PhiPack package (Bruen 2005) with 10,000 permutations per test and a window size of 100 nucleotides. The PhiPack package is also capable of calculating the PHI statistic along a sliding window to determine which portions of an alignment exhibit the strongest signal of mosaicism. An initial alignment profile was computed with default window size (-w = 100), scanning size (-n = 1000), and step size (-m = 25). 3401 tests were performed, and of these 412 recovered too few variable sites within the scanning window to compute a PHI score. Parameters were adjusted such that scanning size was equal to 1500 and window size equal to 50. Step size remained unchanged.

### 3.3 Direct tests for recombination and breakpoint identification

The data were subjected to three direct tests for recombination: SBP for the detection of a single, best-supported breakpoint, GARD for detection of multiple breakpoints, and geneconv for detection of potential gene conversion events. The HyPhy command line package v2.1.1 (Kosakovsky Pond et al. 2005) implements both the SBP and GARD methods (Kosakovsky Pond et al. 2006). Because no systematic testing of SBP or GARD has been performed with non-
default parameters, default parameters were used in this study. SBP and GARD both require a user-specified model for nucleotide substitution. Heirarchical likelihood ratio testing in jModelTest v2.1.7 (Darrida et al. 2012; Guindon and Gascuel 2003) identified the GTR+I+G model as optimal according to BIC and AIC criteria.

The SBP algorithm was used to search for a single best-supported breakpoint according to the AIC criterion with a general reversible (GRM) model of nucleotide substitution and local, independent estimation of all model parameters on each branch. GARD was used to detect multiple breakpoints using a general time reversible (GTR) model with a general discrete distribution of rate variation on 2 bins (coding/non-coding sequence). Lack of recombination in each fragment identified by GARD was confirmed using the PHI test with 10,000 permutations.

The GENECONV package v1.81a (Sawyer 1999) implements Sawyer’s (1989) geneconv test for gene conversion. In this study only global permuted p-values were considered, as these are automatically corrected for multiple sequence comparisons. Data were permuted 10,000 times to compute significance of candidate conversion fragments. Only slight modification of default GENECONV parameters were made such that the circularity of the chloroplast genome was assumed and sites with missing data were omitted. Gapped sites were excluded because no allowance was made for the inclusion of potentially informative INDELs during the sequence alignment stage.

3.4 Phylogenetic analysis

Phylogenetic topology tests for tree incongruence were used to confirm the validity of putative breakpoints identified by GARD. The full 86-Kbp alignment was partitioned into spans identified by GARD as being non-recombinant, i.e. spans bounded on either side by breakpoints
identified in the optimal breakpoint model. The number of phylogenetically informative sites in each span was quantified using Paup* v4.0a146 (Swofford 2003). Paup* was also used to perform Archie-Faith-Cranston randomized permutation (PTP) tests for phylogenetic structure with 100 replicates per partition, retaining 100 trees from each replicate. Every span identified as having significant phylogenetic signal was subjected to Bayesian analysis using MrBayes v3.2.5 (Huelsenbeck and Ronquist 2001; Ronquist and Huelsenbeck 2003; Altekar et al. 2004).

No distinction was made between first, second, and third codon positions and each sequence fragment was treated as a single data partition. Each partition was assigned a GTR+I+G model. State frequencies, substitution rates, transition/transversion ratios, gamma shape parameters, and proportion of invariable sites were estimated empirically for each reduced data set during Markov Chain Monte Carlo (MCMC) runs. All analyses were performed with two independent runs (-nruns = 2) initiated with different starting seeds, with four chains (-nchains = 4) per run. Chains were sampled every 100 generations, and MCMC diagnostics were performed every 1000 generations. Chains were initially run for 10,000 generations, after which chain heating (-temp) was adjusted such that acceptance rates for chain swapping upon repeat analysis ranged between 10% and 70%. Chains were then run for 5,000,000 generations each. The first 25% of trees were discarded as burn-in.24 Runs were considered to have converged when convergence diagnostics output by sump approached 1 and when the average standard deviation of split frequencies was ≤0.01, indicating convergence upon a stationary distribution. Unrooted consensus trees lacking branch lengths were output according to a 50% majority rule (minimum 50% PP on displayed bipartitions) using sumt.

24 Samples discarded at the beginning of an MCMC run before chains have reached convergence
A concatenated set of 18 trees generated from non-recombinant partitions were subjected to topology testing using TREE-PUZZLE v5.3.rc (Schmidt et al. 2002) and CONSEL v0.20 (Shimodaira and Hasegawa 2001). All sequence fragments were compared independently to the concatenated tree set. TREE-PUZZLE was used to write estimated site-log-likelihood values (-wsl) for each sequence fragment with the following non-default parameters: parameter estimation (-x) utilizing a neighbor-joining (NJ) starting tree, a gamma-distributed model of rate heterogeneity (-w) with four rate categories, and a GTR model of nucleotide substitution (-m). Substitution rates empirically estimated by MrBayes during MCMC runs were substituted for default substitution rates of 1.0. Site-log-likelihoods were then input into CONSEL for topology testing using the approximately unbiased (Shimodaira 2002) and Shimodaira-Hasegawa tests (Shimodaira and Hasegawa 1999).

4. Results

4.1 Alignment and genotyping

Sequence length before correction of alignment errors was 86-Kbp. Final alignment length after manual adjustment was 86023-Bp. A total of 23251 sites contained at least one gap (-), and PhiPack, geneconv, PTP, and Bayesian analyses considered these sites missing data. 62772 gap-free sites remained (72.97%) and of these, 61813 (98.47%) were monomorphic. Of 959 variable sites, 584 (60.89%) were autapomorphies and 375 (39.10%) were informative. When sites containing some missing data were considered, 1008 variable sites (1.17% of total alignment length) and 569 informative sites (0.66%) were found.
4.2 Sequence summary statistics and indirect tests for recombination

No haplotypes were shared by any two or more samples in the alignment. Per-site nucleotide diversity ($\pi$) was found to be $1.58e-3$, with a standard deviation of $1.7e-4$. Per-site mutation rate ($\theta$) was $3.16e-3$ with a standard deviation of $8.1e-4$ assuming the absence of recombination and $1.0e-4$ assuming free recombination. The average number of nucleotide differences ($k$) per genotype was $98.98$ with a standard deviation of $43.03$ in the absence of recombination and $5.83$ assuming free recombination.

To estimate the recombination parameter ($R$) and the minimum number of recombination events ($RM$), 959 polymorphic characters were considered. The average nucleotide distance between the most distant sites ($D$) in the alignment was $84636.76$-Bp. The per-gene estimate of the recombination parameter ($R$) was $4.5$. $R$ here is interpreted as the recombination parameter for the entire alignment of 86-Kbp. Estimated $R$ between adjacent sites was $1.0e-4$. Using the estimated parameter $R$, 449826 pairwise comparisons of variable sites were analyzed for signs of recombination using the four-gamete test. Of these, 33490 pairs of sites among 67 individuals were found to have all four gametic types (00, 01, 10, 11), and the most parsimonious number of recombination events ($RM$) estimated to have occurred was 197 (data not shown).

PhiPack likewise recovered a total of 959 unambiguous polymorphic sites excluding gapped sites. Of all pairwise comparisons performed, Max $X^2$ recovered a single best breakpoint at nucleotide position 61035 in a pairwise comparison of *D. obovata* (203) and *I. grandiflorum* (211). Permuted Max $X^2$, NSS, and PHI found very significant evidence of recombination with $p < 0.01$. A PHI profile test with scan size 1500 and window size 50 divided the alignment into 3381 total windows, of which 3280 possessed sufficient informative sites for non-permuted estimation of PHI. The PHI statistic was found to be significant ($p < 0.01$) in 127 individual tests.
distributed across the alignment in 9 non-consecutive mosaic fragments (Figure 1), with these fragments beginning at sites \{26150, 34100, 41075, 42075, 46025, 48925, 53675, 72675, 72825\}.

4.3 Direct tests for recombination and breakpoint identification

SBP analysis found a single breakpoint between position 64,000 and 71,000 with \( p < 0.01 \). GARD recovered 17 breakpoints in the optimal breakpoint model, dividing the full alignment of 86-Kbp into 18 non-recombinant sequence fragments (Figure 2). Average fragment length was 4778-Bp, and each span contained an average of 33 phylogenetically informative characters (PICs) (Table 2). PHI tests confirmed lack of recombinant signal in each sequence fragment, though four fragments (4, 5, 9, 18) retained a signal of recombination according to either the Max \( X^2 \) or NSS tests. It was assumed that these results were due to homoplasy rather than recombination, to which the PHI test is more robust than either of the latter statistics. PTP tests confirmed the presence of phylogenetic structure in all 18 fragments, finding in each case that the most parsimonious tree was significantly shorter than trees derived from permuted data.

GENECONV recovered 7 significant gene conversion events (Table 3, Figure 3) between 7 unique sequence pairs comprised of three source genomes and 7 unique destination genomes. Significance was judged on the basis of global permuted p-values. \textit{D. obovata} (203) was identified as the source sample in 5 of 7 of these events, and GENECONV results independently verified Max \( X^2 \) results by identifying \textit{D. obovata} (203) and \textit{I. grandiflorum} (211) as a conversion pair. Max \( X^2 \) identified position 61035 as the likeliest breakpoint in a pairwise comparison of these two samples, while GENECONV finds the bounds of converted sequence in this pair to be at \(~54300\) and \(~58500\).
4.4 Phylogenetic analyses

The approximately unbiased (AU) and Shimodaira-Hasegawa (SH) test were used to screen for phylogenetic incongruence because they are optimized for 1) tree sets specified a priori, 2) tree sets containing one or more trees inferred from the sequence data in question, and 3) tree sets containing three or more topologies (in the case of AU) (Shimodaira 2002). All three factors render the often-used KH test inappropriate for these data. The SH test is optimized to reduce type-1 error in cases where no more than 2 trees are being compared, but results of the SH test are heavily biased towards conservatism in comparisons of > 2 trees (Shimodaira 2002). The AU test, by comparison, is designed for tests of tree sets containing many trees, and these results have been emphasized in Figure 4.

Due to the paucity of phylogenetically informative sites retained in non-recombinant sequence fragments, all Bayesian analyses recovered poorly resolved trees with few bipartitions supported by > 90% PP (see Supplementary Figures 4-21). Nevertheless, visual inspection of well-supported clades found multiple instances of incongruence between nodes subtending two leaves with ≈90% posterior probability. Comparisons between sequence fragments and 50% majority rule consensus trees using SH and AU topology incongruence tests found apparent widespread evidence for fragment-tree incongruence (but see Discussion).

5. Discussion

All methods used in this study found significant evidence for recombination, though it is uncertain whether these results were due to true reticulation\textsuperscript{25} or to artefactual recombination\textsuperscript{26}.

\textsuperscript{25} Evolutionary change dependent upon recombination
\textsuperscript{26} Patterns suggestive of recombination that are in fact due to sequencing or assembly error
Artefactual recombination may have been introduced during the alignment stage if nuclear reads sharing homology with the chloroplast LSC were retained. An attempt was made to determine the likely extent of contamination, but without thorough realignment of all data the true severity of nuclear contamination cannot be ascertained. Contamination may also have been introduced during the sequencing stage, if genetic material was transferred between wells. Well contamination may result in uneven representation of two or more individuals’ DNA in a sample. During alignment, these contaminated data may create a perfect signal of recombination. By requiring a minimum of 80% of reads support an alternative allele, an attempt was made to reduce error due to sample contamination, but if coverage of the chloroplast by contaminating reads was patchy, blocks of aligned sequence sharing homology with the contaminant haplotype may have been retained.

Assuming sample contamination was corrected by the genotyping pipeline, the consistently significant results found by methods utilized in this study are highly suggestive of a history of reticulation in the chloroplasts of *Iochrominae*, though not all of the specific recombination events detected are biologically plausible. Bayesian phylogenetic analyses of non-recombinant fragments and GENECONV may both provide information about the participants and directionality of recombination events. GENECONV recovered seven candidate gene conversions, but transfer of genetic material between participant species is geographically implausible. It should be noted that species and population sampling was not exhaustive, and individuals belonging to participant species may occur or have once occurred in sympathy in unsampled locales. Of the conversion pairs found, only *I. convertiflorum* and *I. cornifolium* are known to occur in sympathy and share a significant proportion of their pollinator assemblages (30%) (Smith et al. 2008), though natural hybridization between species of *Iochroma*,
*Eriolarynx*, and *Dunalia* is hypothetically possible if some portion of their pollinator assemblages are shared (Smith and Baum 2007).

GARD partitioned full the 86-Kbp alignment into non-recombinant fragments averaging approximately 4-Kbp in length and possessing an average of 33 phylogenetically informative sites. 33 sites would typically not be sufficient to construct a well-supported phylogeny for 67 individuals, and this was found to be the case in this study. With few clades supported by ≥ 90% posterior probability, it was not possible upon visual inspection to find conflicting relationships suggestive of plastid capture, which may be indicative of incomplete lineage sorting or paternal leakage. A handful of well-supported nodes subtending two taxa were found to be incongruent, though it was not possible to ascertain without additional analysis whether these apparent topological conflicts are due to reticulation or homoplasy. If well-supported, conflicting nodes are taken as indicators of recombination, several biologically implausible recombination events have been suggested. For instance, incongruent nodes supporting relationships between *Iochroma ellipticum* and all other species in this data set (e.g. Supplementary Figures 4, 9, and 15), even its closest relatives according to the study of Smith and Baum (2006), are suggestive of hybridization between Andean taxa and a species native to the Galapagos (*I. ellipticum*). It can reasonably be assumed that insufficient phylogenetically informative characters were present in short fragments for biologically reasonable conclusions to be drawn from topologies. Likewise, the widespread, significant phylogenetic conflict recovered by approximately unbiased and Shimodaira-Hasegawa tests for topological incongruence may not be representative of true hard incongruence between sequence fragments. Topology tests did not factor in relative node support, and nodes with less than 90% PP should likely have been collapsed into polytomies.
before analysis. Alternatively, well-supported clades should have been independently confronted with the data.

The strongest indicators of recombinant history in these data are perhaps the three modern, non-parametric indirect tests utilized (NSS, Max $X^2$, and PHI), and the results of SBP. SBP has been demonstrated to be very robust in tests of both simulated and empirical data sets (Kosakovsky Pond et al. 2006). Max $X^2$, SBP, PHI, and phylogenetic analyses all indicate mosaicism in the region between ~60-Kbp and ~70-Kbp, and the consistency of these results suggests the presence of at least one, well-supported recombination event in the LSC of Iochrominae. Max $X^2$ suggested D. obovata (203) and I. grandiflorum (211) as the likeliest participants of a recombination event at approximately 61-Kbp, and although GENECONV likewise finds support for recombination between these two sequences, it places the breakpoint approximately 6.5-Kbp earlier at position 54300. Patterns of phylogenetic incongruence between nodes subtending two taxa suggests consistent conflict between trees generated from early fragments (1-14) and later fragments (15-18), which correspond to positions 1-71186 and 71187-86023 respectively. The PHI alignment profile also suggests the presence of mosaicism at ~54-Kbp and ~72-Kbp, and finds no evidence of recombination between 72-Kbp and the end of the alignment. These three results suggest that evolutionary history is not shared, at the very least, between approximately the first 60-Kbp of the alignment and the last 20-Kbp.

Nucleotide diversity ($\pi$) in these data was 0.00158 or 0.158%, far less than the minimum diversity of 5% considered optimal for detecting recombination (Posada 2002). Only 959 variable sites excluding gapped sites were available for analysis, and of these only 375 or 0.603% of the data set excluding missing sites were phylogenetically informative. Few direct attempts have been made to detect recombination in the chloroplasts of higher plants, and so it is
not known how the particular characteristics of chloroplast DNA including low sequence diversity might impact the efficacy of detection methods. It is possible that, given a high degree of sequence conservation and how rarely detectable recombination might be assumed to take place in plastids, the negative effects of low sequence diversity may be lessened.

The effectiveness of modern methods for detecting recombination, both parametric and non-parametric, depend upon a number of factors including the recency of the recombination event, the extent of homoplasy and the nucleotide diversity of the data, the exhaustiveness of the sampling strategy, and the phylogenetic distance between parental sequences. Because no single recombination strategy performs optimally under all conditions, it is recommended that researchers use multiple methods employing different analytic strategies (Posada et al. 2002). Most modern methods for recombination detection either fail to detect reticulation or produce an excess of false positives in situations where recombination rates are high and sequence diversity is either very low or very high due to recurrent evolution.

Mutation, if it occurs frequently, erases similarity shared between parental and recombinant sequences. The rate of degradation is hastened by repeat recombination events. Chloroplast DNA is structurally and sequentially conserved, and though there is a growing body of evidence suggesting that chloroplast sequence in diverse angiosperm lineages is evolving adaptively, the number of loci showing evidence of adaptive selection is fairly small and evolution is believed to occur over very long time scales (Bock et al. 2014). Sharing of homologous DNA in the form of gene conversion is known to play an important role in the maintenance of chloroplast genomic integrity (Marechal and Brisson 2009), but the frequency of inter-organellar conversion, if it occurs, cannot be inferred from sequence data alone. Recombination between chloroplast genomes, either by horizontal gene transfer or gene
conversion, would likely only be detectable at the inter-organellar level within heteroplasmic cells.

Heteroplasmy has been invoked as a driving force in mtDNA recombination, even in instances where heteroplasmy was not itself detected (Stadler and Delph 2002; Jaramilla-Correa and Bousquet 2005). The persistence of multiple chloroplast haplotypes within individual cells and organisms is known to be widespread, if not frequent, in angiosperms (Wagner et al. 1991; Barr et al. 2005; Ellis et al. 2008; Pearl et al. 2009; Bentley et al. 2010; Nunes et al. 2013). Heteroplasmy may arise either as a consequence of point mutation or paternal leakage, which has been demonstrated in up to 30% of angiosperm lineages and may be ubiquitous in higher plants practicing uniparental maternal inheritance of cytoplasm (Medgyesy et al. 1986; Schmitz and Kowallik 1986; Cornu and Dulieu 1988; Horlow et al. 1990; Sewell et al. 1993; Azhagiri and Maliga 2007).

Demonstrating heteroplasmy in chloroplasts may not be possible through next-gen sequencing and alignment procedures, but paternal leakage may be suggested by discordance between cytoplasmic and nuclear loci where topologies are sufficiently informative. In recombinant organelles, polymorphic sites sufficient for phylogenetic analysis may not be present in non-recombinant fragments, particularly if recombination has occurred between relatively closely related species. If divergence between parental sequences occurred recently, their slowly evolving plastids may not have had adequate time to accumulate mutations sufficient to differentiate them. And yet, if the goal is to detect homologous chloroplast recombination within heteroplasmic cells, divergence of parental species must have occurred recently enough for interspecific hybridization to be possible. Minimum sequence diversity of 5% is recommended for successful detection of recombination by most methods (Posada 2002),
but this amount of diversity may not be found within a chloroplast data set except among very distantly related species that could not be reasonably expected to be capable of interbreeding. Intergeneric hybridization among *Iochrominae*’s six traditional genera is possible, and under greenhouse condition hybridization is known to result in the production of viable F1 offspring. It may be suggested that even if only 2% of hybrid offspring inherit both paternal and maternal cytoplasm as was observed by Ellis et al. (2008) in hybrid *Helianthus, Iochrominae* hybrids may present a rare opportunity to investigate infrequent reticulation in the chloroplasts of higher plants.

6. Conclusions and Future Research

Future analysis of these data may incorporate tests to gauge the prevalence of homoplasy in *Iochrominae* plastids, and to tease apart which homoplasious traits can be attributed to recombination and to other biological processes such as convergent and parallel evolution. Results presented here indicate that phylogenetic studies incorporating chloroplast DNA should make an attempt prior to analysis to ascertain the extent of heteroplasmy and recombination in the taxa to be analyzed. In addition to compromising phylogenetic reconstruction, recombination of chloroplast DNA has far-reaching implications for agricultural and phylogeographic research. A growing body of evidence suggests that the evolutionary histories of mitochondrial DNA may not be as simple as once thought, and in light of these findings, reevaluation of traditional assumptions about the tree-like histories of chloroplast DNA is warranted.
<table>
<thead>
<tr>
<th>Spp.</th>
<th>ID #</th>
<th>Dash #</th>
<th>HP</th>
<th>Parent (H) / Offspring (P)</th>
<th>Clade</th>
<th>Samp. Loc.</th>
<th>Lat (°S)</th>
<th>Long (°W)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Acnistus arborescens</td>
<td>98</td>
<td>334</td>
<td>P</td>
<td>A. arb x I. cy, I. smith, I. stenanthum, I. tupayachianum</td>
<td>A</td>
<td>Peru</td>
<td>7.42409</td>
<td>78.90111</td>
</tr>
<tr>
<td>Acnistus arborescens</td>
<td>192</td>
<td>2686</td>
<td>P</td>
<td>A. arb x I. cy, I. smith, I. stenanthum, I. tupayachianum</td>
<td>A</td>
<td>NA</td>
<td>NA</td>
<td>NA</td>
</tr>
<tr>
<td>Acnistus arborescens</td>
<td>250</td>
<td>747</td>
<td>P</td>
<td>A. arb x I. cy, I. smith, I. stenanthum, I. tupayachianum</td>
<td>A</td>
<td>NA</td>
<td>NA</td>
<td>NA</td>
</tr>
<tr>
<td>Acnistus arborescens</td>
<td>2428</td>
<td>1276</td>
<td>P</td>
<td>A. arb x I. cy, I. smith, I. stenanthum, I. tupayachianum</td>
<td>A</td>
<td>Greenhouse</td>
<td>NA</td>
<td>NA</td>
</tr>
<tr>
<td>A. arborescens x I. cyaneum</td>
<td>F1</td>
<td>2</td>
<td>H</td>
<td>A. arborescens x I. cyaneum</td>
<td>-</td>
<td>Greenhouse</td>
<td>NA</td>
<td>NA</td>
</tr>
<tr>
<td>Dunalia brachyacantha</td>
<td>188</td>
<td>506</td>
<td>-</td>
<td>-</td>
<td>NA</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>Dunalia brachyacantha</td>
<td>208</td>
<td>359</td>
<td>-</td>
<td>-</td>
<td>Bolivia</td>
<td>18.10605</td>
<td>64.04406</td>
<td></td>
</tr>
<tr>
<td>Dunalia obovata</td>
<td>203</td>
<td>184</td>
<td>-</td>
<td>-</td>
<td>Peru</td>
<td>11.68201</td>
<td>75.56978</td>
<td></td>
</tr>
<tr>
<td>Dunalia solanaceae</td>
<td>185</td>
<td>930</td>
<td>-</td>
<td>-</td>
<td>Ecuador</td>
<td>0.2968</td>
<td>78.65029</td>
<td></td>
</tr>
<tr>
<td>Dunalia solanaceae</td>
<td>227</td>
<td>1066</td>
<td>-</td>
<td>-</td>
<td>NA</td>
<td>NA</td>
<td>NA</td>
<td>NA</td>
</tr>
<tr>
<td>Dunalia solanaceae</td>
<td>230</td>
<td>4568</td>
<td>-</td>
<td>-</td>
<td>Ecuador</td>
<td>0.9563</td>
<td>78.98035</td>
<td></td>
</tr>
<tr>
<td>Dunalia spathulata</td>
<td>136</td>
<td>1367</td>
<td>-</td>
<td>-</td>
<td>Peru</td>
<td>9.83831</td>
<td>76.11503</td>
<td></td>
</tr>
<tr>
<td>Dunalia spathulata</td>
<td>141</td>
<td>2832</td>
<td>-</td>
<td>-</td>
<td>Peru</td>
<td>9.84009</td>
<td>76.12047</td>
<td></td>
</tr>
<tr>
<td>Dunalia spinosa</td>
<td>114</td>
<td>300</td>
<td>-</td>
<td>-</td>
<td>NA</td>
<td>NA</td>
<td>NA</td>
<td>NA</td>
</tr>
<tr>
<td>Eriolarynx fasciculata</td>
<td>133</td>
<td>1711</td>
<td>-</td>
<td>-</td>
<td>Bolivia</td>
<td>17.46477</td>
<td>65.75217</td>
<td></td>
</tr>
<tr>
<td>Eriolarynx ochromoides</td>
<td>238</td>
<td>2686</td>
<td>-</td>
<td>-</td>
<td>Argentina</td>
<td>27.40167</td>
<td>65.97885</td>
<td></td>
</tr>
<tr>
<td>Eriolarynx loenzii</td>
<td>28</td>
<td>1418</td>
<td>-</td>
<td>-</td>
<td>E</td>
<td>NA</td>
<td>NA</td>
<td>NA</td>
</tr>
<tr>
<td>Eriolarynx loenzii</td>
<td>184</td>
<td>202</td>
<td>-</td>
<td>-</td>
<td>E</td>
<td>Bolivia</td>
<td>27.40198</td>
<td>65.97917</td>
</tr>
<tr>
<td>Iochroma albi</td>
<td>174</td>
<td>1049</td>
<td>-</td>
<td>-</td>
<td>Ecuador</td>
<td>4.64619</td>
<td>79.72258</td>
<td></td>
</tr>
<tr>
<td>Iochroma australis</td>
<td>117</td>
<td>239</td>
<td>-</td>
<td>-</td>
<td>E</td>
<td>Bolivia</td>
<td>20.78464</td>
<td>65.04123</td>
</tr>
<tr>
<td>Iochroma australis</td>
<td>183</td>
<td>357</td>
<td>-</td>
<td>-</td>
<td>E</td>
<td>Argentina</td>
<td>27.66603</td>
<td>65.94261</td>
</tr>
<tr>
<td>Iochroma ayabacense</td>
<td>126</td>
<td>246</td>
<td>H</td>
<td>I. cyaneum x I. lehmani squamosum</td>
<td>-</td>
<td>NA</td>
<td>NA</td>
<td>NA</td>
</tr>
<tr>
<td>Iochroma ayabacense</td>
<td>248</td>
<td>449</td>
<td>H</td>
<td>I. cyaneum x I. lehmani squamosum</td>
<td>-</td>
<td>NA</td>
<td>NA</td>
<td>NA</td>
</tr>
<tr>
<td>Iochroma ayabacense</td>
<td>105</td>
<td>3013</td>
<td>H</td>
<td>I. confertiflorum x I. cyaneum</td>
<td>-</td>
<td>Ecuador</td>
<td>4.619405</td>
<td>79.71061</td>
</tr>
<tr>
<td>Iochroma baumii</td>
<td>180</td>
<td>2145</td>
<td>-</td>
<td>-</td>
<td>Ecuador</td>
<td>0.3764</td>
<td>78.12299</td>
<td></td>
</tr>
<tr>
<td>Iochroma calycinum</td>
<td>229</td>
<td>468</td>
<td>-</td>
<td>-</td>
<td>F</td>
<td>Ecuador</td>
<td>0.95636</td>
<td>78.98084</td>
</tr>
<tr>
<td>Iochroma confertiflorum</td>
<td>171</td>
<td>3065</td>
<td>P</td>
<td>I. tupayachianum</td>
<td>A</td>
<td>Ecuador</td>
<td>4.68037</td>
<td>79.77975</td>
</tr>
<tr>
<td>Iochroma confertiflorum</td>
<td>226</td>
<td>1134</td>
<td>P</td>
<td>I. tupayachianum</td>
<td>A</td>
<td>Ecuador</td>
<td>1.90748</td>
<td>78.98524</td>
</tr>
<tr>
<td>Iochroma cornifolium</td>
<td>187</td>
<td>1002</td>
<td>P</td>
<td>I. stenanthum</td>
<td>C</td>
<td>Peru</td>
<td>7.39903</td>
<td>78.90137</td>
</tr>
<tr>
<td>Iochroma cornifolium</td>
<td>231</td>
<td>2476</td>
<td>P</td>
<td>I. stenanthum</td>
<td>C</td>
<td>Ecuador</td>
<td>2.23975</td>
<td>78.81111</td>
</tr>
<tr>
<td>Iochroma cornifolium</td>
<td>232</td>
<td>1536</td>
<td>P</td>
<td>I. stenanthum</td>
<td>C</td>
<td>Ecuador</td>
<td>2.23069</td>
<td>78.81594</td>
</tr>
<tr>
<td>Iochroma cornifolium</td>
<td>249</td>
<td>643</td>
<td>P</td>
<td>I. stenanthum</td>
<td>C</td>
<td>NA</td>
<td>NA</td>
<td>NA</td>
</tr>
<tr>
<td>Iochroma cyaneum</td>
<td>265</td>
<td>1526</td>
<td>P</td>
<td>A. arb x I. cy, I. ayabacense, I. ayabacense pingola</td>
<td>C</td>
<td>Greenhouse</td>
<td>NA</td>
<td>NA</td>
</tr>
<tr>
<td>I. cyaneum x A. arborescens</td>
<td>191</td>
<td>911</td>
<td>H</td>
<td>A. arborescens x I. cyaneum</td>
<td>-</td>
<td>Greenhouse</td>
<td>NA</td>
<td>NA</td>
</tr>
<tr>
<td>I. cyaneum x A. arborescens</td>
<td>247</td>
<td>1288</td>
<td>H</td>
<td>A. arborescens x I. cyaneum</td>
<td>-</td>
<td>NA</td>
<td>NA</td>
<td>NA</td>
</tr>
<tr>
<td>Iochroma edule</td>
<td>173</td>
<td>688</td>
<td>-</td>
<td>A</td>
<td>Peru</td>
<td>7.94837</td>
<td>78.56065</td>
<td></td>
</tr>
<tr>
<td>Iochroma edule</td>
<td>219</td>
<td>5558</td>
<td>-</td>
<td>A</td>
<td>Peru</td>
<td>7.928183</td>
<td>78.583683</td>
<td></td>
</tr>
<tr>
<td>Iochroma ellipticum</td>
<td>216</td>
<td>4148</td>
<td>-</td>
<td>A</td>
<td>Greenhouse</td>
<td>NA</td>
<td>NA</td>
<td>NA</td>
</tr>
<tr>
<td>Iochroma ellipticum</td>
<td>225</td>
<td>1976</td>
<td>-</td>
<td>A</td>
<td>Greenhouse</td>
<td>NA</td>
<td>NA</td>
<td>NA</td>
</tr>
<tr>
<td>Iochroma geosneroides</td>
<td>201</td>
<td>3137</td>
<td>-</td>
<td>F</td>
<td>Ecuador</td>
<td>0.0214</td>
<td>78.55106</td>
<td></td>
</tr>
<tr>
<td>Iochroma grandiflorum</td>
<td>142</td>
<td>411</td>
<td>-</td>
<td>U</td>
<td>Peru</td>
<td>7.38543</td>
<td>78.89542</td>
<td></td>
</tr>
<tr>
<td>Iochroma grandiflorum</td>
<td>211</td>
<td>384</td>
<td>-</td>
<td>U</td>
<td>Peru</td>
<td>7.36186</td>
<td>78.88052</td>
<td></td>
</tr>
</tbody>
</table>
Table 1. Taxa present in the data set are listed according to species or ecotype. In tree format (Supplementary 4-21) species names and sample ID numbers are concatenated. Dash counts are equivalent to the number of sites per individual with missing data. Species status as a known or probable hybrid or parental taxa is given, if known (S. D. Smith, personal correspondence, October 2014). Where taxa are suspected hybrids, putative parental taxa are listed in the adjacent column and vice versa. If taxa have been assigned to a monophyletic clade in the revised classification system of Smith and Baum (2006), these classifications are indicated (see Supplementary Figure 1). Daniel Gates at the University of Lincoln-Nebraska provided sampling locations and coordinates.
Figure 1. Significance of pairwise homoplasy indices across the LSC

Filled circles along the x-axis correspond to the first site in non-consecutive scanning windows wherein PHI detected a significant signal of recombination with $p < 0.01$. The dotted line indicates a significance cutoff of $\alpha = 0.01$.
Figure 2. Stepwise GARD optimization of a model for optimal \( B \) breakpoints reached stationarity at \( B = 17 \). Stationarity was declared when \( \Delta \text{AIC}_c \) values remained unchanged for 100 generations. The 86-Kbp alignment was partitioned into non-recombinant fragments bounded on either size by breakpoints corresponding to nucleotide positions given in the optimal model (BP = 17, \( \text{AIC}_c = 270672 \)).
Table 2. Summary of non-recombinant fragment statistics

<table>
<thead>
<tr>
<th>Fragment</th>
<th>Start (LSC)</th>
<th>End (LSC)</th>
<th>Length (Frag)</th>
<th>PIC</th>
<th>PTP (p-value)</th>
<th>PHI (p-value)</th>
<th>Max X^2 (p-value)</th>
<th>NSS (p-value)</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>1</td>
<td>5535</td>
<td>5534</td>
<td>39</td>
<td>0.010 **</td>
<td>0.819</td>
<td>0.623</td>
<td>0.291</td>
</tr>
<tr>
<td>2</td>
<td>5536</td>
<td>12877</td>
<td>7341</td>
<td>37</td>
<td>0.010 **</td>
<td>0.970</td>
<td>0.406</td>
<td>0.987</td>
</tr>
<tr>
<td>3</td>
<td>12878</td>
<td>18005</td>
<td>5127</td>
<td>28</td>
<td>0.010 **</td>
<td>0.892</td>
<td>0.216</td>
<td>0.468</td>
</tr>
<tr>
<td>4</td>
<td>18006</td>
<td>27675</td>
<td>9669</td>
<td>38</td>
<td>0.010 **</td>
<td>0.919</td>
<td>0.216</td>
<td>0.012 *</td>
</tr>
<tr>
<td>5</td>
<td>27676</td>
<td>31129</td>
<td>3453</td>
<td>25</td>
<td>0.010 **</td>
<td>0.403</td>
<td>0.133</td>
<td>0.003 **</td>
</tr>
<tr>
<td>6</td>
<td>31130</td>
<td>36206</td>
<td>5076</td>
<td>20</td>
<td>0.010 **</td>
<td>0.812</td>
<td>0.106</td>
<td>0.314</td>
</tr>
<tr>
<td>7</td>
<td>36207</td>
<td>45022</td>
<td>8815</td>
<td>38</td>
<td>0.010 **</td>
<td>0.389</td>
<td>0.038</td>
<td>0.417</td>
</tr>
<tr>
<td>8</td>
<td>45023</td>
<td>49913</td>
<td>4890</td>
<td>42</td>
<td>0.010 **</td>
<td>0.162</td>
<td>0.306</td>
<td>0.120</td>
</tr>
<tr>
<td>9</td>
<td>49914</td>
<td>52306</td>
<td>2392</td>
<td>20</td>
<td>0.010 **</td>
<td>0.671</td>
<td>0.000 **</td>
<td>0.334</td>
</tr>
<tr>
<td>10</td>
<td>52307</td>
<td>56567</td>
<td>4260</td>
<td>23</td>
<td>0.010 **</td>
<td>0.703</td>
<td>0.476</td>
<td>0.682</td>
</tr>
<tr>
<td>11</td>
<td>56568</td>
<td>59842</td>
<td>3274</td>
<td>22</td>
<td>0.010 **</td>
<td>0.606</td>
<td>0.330</td>
<td>0.377</td>
</tr>
<tr>
<td>12</td>
<td>59843</td>
<td>64069</td>
<td>4226</td>
<td>25</td>
<td>0.010 **</td>
<td>0.197</td>
<td>0.393</td>
<td>0.717</td>
</tr>
<tr>
<td>13</td>
<td>64070</td>
<td>66082</td>
<td>2012</td>
<td>23</td>
<td>0.010 **</td>
<td>0.350</td>
<td>0.504</td>
<td>0.132</td>
</tr>
<tr>
<td>14</td>
<td>66083</td>
<td>71186</td>
<td>5103</td>
<td>32</td>
<td>0.010 **</td>
<td>0.781</td>
<td>0.367</td>
<td>0.030</td>
</tr>
<tr>
<td>15</td>
<td>71187</td>
<td>73040</td>
<td>1853</td>
<td>33</td>
<td>0.010 **</td>
<td>0.060</td>
<td>0.630</td>
<td>0.813</td>
</tr>
<tr>
<td>16</td>
<td>73041</td>
<td>76142</td>
<td>3101</td>
<td>54</td>
<td>0.010 **</td>
<td>0.538</td>
<td>0.480</td>
<td>0.967</td>
</tr>
<tr>
<td>17</td>
<td>76143</td>
<td>82422</td>
<td>6279</td>
<td>54</td>
<td>0.010 **</td>
<td>0.090</td>
<td>0.071</td>
<td>0.664</td>
</tr>
<tr>
<td>18</td>
<td>82423</td>
<td>86023</td>
<td>3600</td>
<td>41</td>
<td>0.010 **</td>
<td>0.407</td>
<td>0.013 *</td>
<td>0.976</td>
</tr>
</tbody>
</table>

Table 2. Location, length, number of phylogenetically informative sites (PIC), permutation test (PTP) p-values, and results of three indirect tests for recombination for each of 18 non-recombinant fragments are shown. All fragments exhibited significant evidence of phylogenetic structure according to PTP tests, and no fragment contained site pairs with significant refined incompatibility according to the PHI statistic.
Table 3. Results of GENECONV analysis of gene conversion

<table>
<thead>
<tr>
<th>Source</th>
<th>Destination</th>
<th>Location (S)</th>
<th>Location (D)</th>
<th>Length (bp)</th>
<th>P-value</th>
<th>P-value (BC)</th>
<th>Polymorphisms</th>
<th>Distance (km)</th>
</tr>
</thead>
<tbody>
<tr>
<td>D. obovata (203)</td>
<td>I. grandiflorum (211)</td>
<td>54345</td>
<td>54257</td>
<td>4216</td>
<td>0.0026 **</td>
<td>0.0138 *</td>
<td>32</td>
<td>1160</td>
</tr>
<tr>
<td>D. obovata (203)</td>
<td>I. umbellatum (127)</td>
<td>54667</td>
<td>54473</td>
<td>5025</td>
<td>0.0072 **</td>
<td>0.0385 *</td>
<td>38</td>
<td>2205</td>
</tr>
<tr>
<td>D. obovata (203)</td>
<td>I. ayabacense (248)</td>
<td>54667</td>
<td>54474</td>
<td>3894</td>
<td>0.0130 *</td>
<td>0.0623</td>
<td>30</td>
<td>NA</td>
</tr>
<tr>
<td>D. obovata (203)</td>
<td>I. tingoense (112)</td>
<td>54667</td>
<td>54634</td>
<td>3894</td>
<td>0.0257 *</td>
<td>0.1290</td>
<td>30</td>
<td>1257</td>
</tr>
<tr>
<td>D. obovata (203)</td>
<td>I. grandiflorum (142)</td>
<td>54667</td>
<td>54542</td>
<td>5031</td>
<td>0.0386 *</td>
<td>0.1830</td>
<td>38</td>
<td>1160</td>
</tr>
<tr>
<td>E. lorentzii (28)</td>
<td>I. baumii (180)</td>
<td>73373</td>
<td>71650</td>
<td>5348</td>
<td>0.0057 **</td>
<td>0.0303 *</td>
<td>44</td>
<td>2080</td>
</tr>
<tr>
<td>I. confertiflorum (171)</td>
<td>I. cornifolium (232)</td>
<td>42788</td>
<td>43548</td>
<td>11636</td>
<td>0.0033 **</td>
<td>0.0179 *</td>
<td>89</td>
<td>482</td>
</tr>
</tbody>
</table>

Table 3. Seven significant gene conversion events were found. Significance values are shown here in the form of permuted p-values with and without Bonferroni correction for multiple comparisons. Locations are sites at which converted fragments begin in (S)ource and (D)estination sequences. Distance in this case is the straight-line geographical distance between the sampling locations of candidate conversion participants in kilometers.
Figure 3. Gene conversion events relative to the *Physalis peruvianum* LSC. Source sequences are indicated by line color (see figure legend), and destination sequences are indicated by line annotations.
Figure 4. Pairwise phylogenetic incongruence between non-recombinant sequence fragments and topologies derived from sequence fragments (see Supplementary Figures 4-21). Shimodaira-Hasegawa (SH) p-values are shown in cell annotations only where they conflict with p-values given by AU tests. In all instances of disagreement, SH p-values were found to be more conservative than AU results as anticipated by Shimodaira (2002)
Acknowledgments

Dr. Nolan Kane and the Kane Lab at CU
   Thesis Advisor, Department of Ecology and Evolutionary Biology
Dr. Barbara Demmig-Adams
   Honors Committee, Department of Ecology and Evolutionary Biology
Dr. Robin Dowell
   Honors Committee, Department of Molecular, Cellular, and Developmental Biology
Dr. Stacey DeWitt Smith and the Smith Lab at UNL
   Assistant Professor, Department of Ecology and Evolutionary Biology
   Assistance with Systematics and the Phylogenetics of Iochrominae
   Sequencing thanks to Daniel Gates, Ph.D. Student, Smith Lab at UNL
Kyle Keepers
   Ph.D. Student, Martin Lab, Department of Ecology and Evolutionary Biology
   Genotyping Pipeline
Silas Tittes
   Ph.D. Student, Kane Lab, Department of Ecology and Evolutionary Biology
   Genotyping Pipeline
The EBIO Honors Program at CU
References


Jakobsen I, Easteal S. 1996. A program for calculating and displaying compatibility matrices as an aid in...


White D, Bryant D, Gemmell N. 2013. How good are indirect tests at detecting recombination in human mtDNA? G3 (Bethesda) 3(7): 1095-1104.


Supplementary Material

Supplementary Figure 1. Phylogenetics of *Iochrominae* according to three nuclear loci (Smith and Baum 2006)
Supplementary Figure 2. Alignment coverage statistics

**Global Median Read Depth**

- min=15
- max=486
- median=68
Supplementary Figure 3. *Physalis peruvianum* reference genome and SNP density map for 67 taxa

Supplementary Figure 3. Locations of single nucleotide polymorphisms (SNPs) (outer) relative to the *Physalis peruvianum* chloroplast genome. Very few polymorphic sites were recovered in the inverted repeat region.
Supplementary Figure 4. Clades with $\geq 50\%$ PP are shown for GARD Fragment 1 (positions 1-5535). Instances of hard incongruence (sister taxa shown non-sister in one or more other trees with $\approx 90\%$ PP) between nodes subtending two taxa are annotated with arrows. For example, the note “211 + 142: 3, 12 - 16” indicates that the clade comprised of *I. grandiflorum* (211) + *I. grandiflorum* (142) is incompatible with placement of *both* (211) and (142) with $\approx 90\%$ PP in trees 3, 12, 13, 14, 15, and 16.
**Supplementary Figure 5.** Clades with ≥ 50% PP are shown for GARD Fragment 2 (positions 5536-12877).
Supplementary Figure 6. Clades with ≥ 50% PP are shown for GARD Fragment 3 (positions 12878-18005).
Supplementary Figure 7. Clades with ≥ 50% PP are shown for GARD Fragment 4 (positions 18006-27675).
**Supplementary Figure 8.** Clades with $\geq 50\%$ PP are shown for GARD Fragment 5 (positions 27676-31129).
Supplementary Figure 9. Clades with ≥ 50% PP are shown for GARD Fragment 6 (positions 31130-36206).
Supplementary Figure 10. Clades with \( \geq 50\% \) PP are shown for GARD Fragment 7 (positions 36207-45022).
Supplementary Figure 11. Clades with ≥ 50% PP are shown for GARD Fragment 8 (positions 45023-49913).
Supplementary Figure 12. Clades with ≥50% PP are shown for GARD Fragment 9 (positions 49914-52306).
**Supplementary Figure 13.** Clades with \( \geq 50\% \) PP are shown for GARD Fragment 10 (positions 52307-56567).
Supplementary Figure 14. Clades with $\geq 50\%$ PP are shown for GARD Fragment 11 (positions 56568–59842).
Supplementary Figure 15. Clades with ≥ 50% PP are shown for GARD Fragment 12 (positions 59843-64069).
Supplementary Figure 16. Clades with ≥ 50% PP are shown for GARD Fragment 13 (positions 64069-66082).
Supplementary Figure 17. Clades with ≥ 50% PP are shown for GARD Fragment 14 (positions 66083-71186).
Supplementary Figure 18. Clades with ≥ 50% PP are shown for GARD Fragment 15 (positions 71187-73040).
Supplementary Figure 19. Clades with ≥ 50% PP are shown for GARD Fragment 16 (positions 73041-76142).
Supplementary Figure 20. Clades with $\geq 50\%$ PP are shown for GARD Fragment 17 (positions 76143-82422).
Supplementary Figure 21. Clades with \( \geq 50\% \) PP are shown for GARD Fragment 18 (positions 82422-86023).