Title

Extensive gene tree discordance and hemiplasy shaped the genomes of North American columnar cacti

Journal

Proceedings of the National Academy of Sciences of the United States of America, 114(45)

ISSN

0027-8424

Authors

Copetti, Dario
Búrquez, Alberto
Bustamante, Enriquena
et al.

Publication Date

2017-11-07

DOI

10.1073/pnas.1706367114

Peer reviewed

Extensive gene tree discordance and hemiplasy shaped the genomes of North American columnar cacti

Dario Copetti[a,b], Alberto Búrquez[c], Enriquena Bustamante[c], Joseph L. M. Charboneau[d], Kevin L. Childs[e], Luis E. Eguiarte[f], Seunghee Lee[a], Tiffany L. Liu[e], Michelle M. McMahon[g], Noah K. Whiteman[h], Rod A. Wing[a,b], Martin F. Wojciechowski[i], and Michael J. Sanderson[d,1]

[a]Arizona Genomics Institute, School of Plant Sciences, University of Arizona, Tucson, AZ 85721; [b]International Rice Research Institute, Los Baños, Laguna, Philippines; [c]Instituto de Ecología, Unidad Hermosillo, Universidad Nacional Autónoma de México, Hermosillo, Sonora, Mexico; [d]Department of Ecology and Evolutionary Biology, University of Arizona, Tucson, AZ 85721; [e]Department of Plant Biology, Michigan State University, East Lansing, MI 48824; [f]Departamento de Ecología Evolutiva, Instituto de Ecología, Universidad Nacional Autónoma de México, Ciudad de México, Mexico; [g]School of Plant Sciences, University of Arizona, Tucson, AZ 85721; [h]Department of Integrative Biology, University of California, Berkeley, CA 94720; and [i]School of Life Sciences, Arizona State University, Tempe, AZ 85287

Edited by David M. Hillis, The University of Texas at Austin, Austin, TX, and approved September 26, 2017 (received for review April 24, 2017)

Few clades of plants have proven as difficult to classify as cacti. One explanation may be an unusually high level of convergent and parallel evolution (homoplasy). To evaluate support for this phylogenetic hypothesis at the molecular level, we sequenced the genomes of four cacti in the especially problematic tribe Pachycereeae, which contains most of the large columnar cacti of Mexico and adjacent areas, including the iconic saguaro cactus (*Carnegiea gigantea*) of the Sonoran Desert. We assembled a high-coverage draft genome for saguaro and lower coverage genomes for three other genera of tribe Pachycereeae (*Pachycereus*, *Lophocereus*, and *Stenocereus*) and a more distant outgroup cactus, *Pereskia*. We used these to construct 4,436 orthologous gene alignments. Species tree inference consistently returned the same phylogeny, but gene tree discordance was high: 37% of gene trees having at least 90% bootstrap support conflicted with the species tree. Evidently, discordance is a product of long generation times and moderately large effective population sizes, leading to extensive incomplete lineage sorting (ILS). In the best supported gene trees, 58% of apparent homoplasy at amino sites in the species tree is due to gene tree-species tree discordance rather than parallel substitutions in the gene trees themselves, a phenomenon termed "hemiplasy." The high rate of genomic hemiplasy may contribute to apparent parallelisms in phenotypic traits, which could confound understanding of species relationships and character evolution in cacti.

Saguaro | homoplasy | lineage sorting | phylogenomics

Cactaceae have undergone adaptive radiation on a continental scale in the Americas. Occurring from arid deserts to alpine steppes and tropical forests, they exhibit a remarkable diversity of growth forms, ranging from tiny, nearly subterranean "buttons" to giant columnar or candelabra forms, epiphytes, and leafy shrubs (1). Classification of the family's 1,438 species (2) has been unusually fraught, with taxonomic treatments recognizing anywhere from 20 to 233 genera (1–4), and classification above the genus level being equally problematic (1, 5). In part this has been attributed to homoplasy (ref. 1, p. 18 and ref. 6, p. 45), the independent evolutionary origin of the same trait (7). For example, early taxonomists combined the giant columnar cacti of North and South America into one large genus *Cereus* Mill (3). Later split into numerous smaller genera, this association was sometimes maintained at a suprageneric rank (8), but molecular phylogenies clearly show North and South American *Cereus* are separate clades (6, 9).

Frequent convergence in growth habit may reflect design limitation (10) of the relatively simple cactus body plan of stem succulence, simplified branching, and loss or reduction of leaves (9) (ref. 11, p. 536). In cacti, convergent simplification via paedomorphosis (12), along with parallelisms among close relatives, has also obscured phylogeny (13, 14). This may help explain the finding that <50% of (nonmonotypic) cactus genera are monophyletic (15). Taxonomic impediments take on special significance in cacti because of the unusually high fraction of species of conservation concern (31% estimated in ref. 16). Despite its potential significance, homoplasy has been quantified in cacti for only a small number of phenotypic traits (14, 17). Here, we focus on the genomes of cacti and assess the degree to which apparent homoplasy among these species is elevated due to discordance between gene trees and the species tree.

An important but recently recognized contributor to molecular homoplasy is "hemiplasy" (18, 19), originally defined as the apparent multiple origin of a character state on the inferred species tree arising when an inferred gene tree (with no homoplasy) is discordant with that species tree (18–21) (Fig. 1). A slight generalization of this definition is needed to account for characters with homoplasy on both trees (Fig. 1, *Lower Right*): A character (site) exhibits hemiplasy if the inferred number of character state changes on the species tree is strictly greater than on the gene tree, which occurs only if the two trees are discordant (*Materials and Methods*).

Significance

Convergent and parallel evolution (homoplasy) is widespread in the tree of life and can obscure evidence about phylogenetic relationships. Homoplasy can be elevated in genomes because individual loci may have independent evolutionary histories different from the species history. We sequenced the genomes of five cacti, including the iconic saguaro of the Sonoran Desert and three other columnar cacti, to investigate whether previously uncharacterized features of genome evolution might explain long-standing challenges to understanding cactus phylogeny. We found that 60% of the amino acid sites in proteins exhibiting homoplasy do so because of conflicts between gene genealogies and species histories. This phenomenon, termed hemiplasy, is likely a consequence of the unusually long generation time of these cacti.

Author contributions: D.C., A.B., M.M.M., N.K.W., R.A.W., M.F.W., and M.J.S. designed research; D.C., A.B., E.B., J.L.M.C., M.M.M., M.F.W., and M.J.S. performed research; D.C., K.L.C., L.E.E., S.L., T.L.L., M.M.M., M.F.W., and M.J.S. analyzed data; and D.C., A.B., E.B., L.E.E., T.L., M.M.M., N.K.W., M.F.W., and M.J.S. wrote the paper.

The authors declare no conflict of interest.

This article is a PNAS Direct Submission.

Data deposition: The sequences reported in this paper have been deposited in GenBank under BioProject number PRJNA318822; individual accession numbers are listed in Tables S1, S3, and S5.

[1]To whom correspondence should be addressed. Email: sanderm@email.arizona.edu.

This article contains supporting information online at www.pnas.org/lookup/suppl/doi:10.1073/pnas.1706367114/-/DCSupplemental.

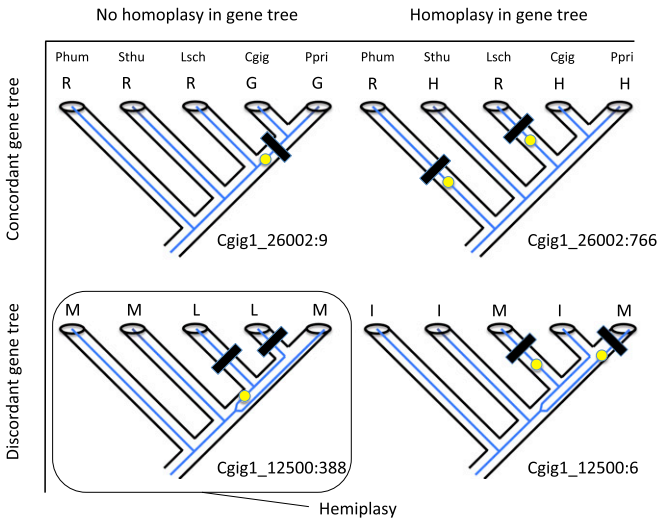

Fig. 1. Hemiplasy in amino acid alignments of cacti (genes annotated as Cgig1_gene##:site ##). Hemiplasy occurs when the inferred number of state changes of a trait is greater on the species tree than on the gene tree, which happens only if the two trees are discordant. Gene trees in blue are imbedded in black species trees. Yellow dot is the position of a state change inferred for an amino acid site for this gene tree. Black rectangles are locations where state changes would be inferred on the species tree (alternative equally optimal reconstructions are possible but the number of state changes is the same). Amino acids are indicated at leaf nodes. Three trees have the same number of state changes on species and gene trees, but at *Lower Left* a homology (one change) on the gene tree is seen as homoplasy (two changes) on the species tree because of discordance (i.e., hemiplasy). Cgig, *C. gigantea*; Lsch, *L. schottii*; Phum, *P humboldtii*; Ppri, *P. pringlei*; Sthu, *S. thurberi*.

Genome scale datasets revealed high levels of gene tree discordance (20, 22–24), which suggests a potentially important role for hemiplasy as a contributor to overall molecular homoplasy. Discordance can arise because of incomplete lineage sorting (ILS) and gene flow, which depend, in turn, on demographic factors at the population level, phylogenetic history, divergence times, mutation rate, generation time, and the timing and taxa involved in introgression or hybridization (25). It can also arise from gene duplication and loss (26).

The cactus tribe Pachycereeae is a model for taxonomic and phylogenetic complexities in cacti (27, 28). Its ∼70 species, including most of the columnar cacti of Mexico and adjacent regions, have been dispersed among 6–23 genera in various taxonomic treatments (1, 5, 27). Despite early molecular support for its monophyly (11), broadly sampled recent molecular studies have led to abandonment of the tribe (9, 13, 15), or to recasting it with the informal name, "core Pachycereeae" (6). Notably, homoplasy in vegetative and floral traits has been cited as a factor contributing to its difficult taxonomic history (ref. 28, p. 1086 and ref. 29, p. 556).

To sample widely in Pachycereeae, we sequenced the genomes of four representatives of the two subtribes of Pachycereeae recognized in Gibson and Horak (30) (Table S1) and some earlier treatments. This included the iconic saguaro cactus of the Sonoran Desert (*Carnegiea gigantea*) to serve as a high-coverage reference assembly, and *Pachycereus pringlei* (cardón, sahueso), *Lophocereus schottii* (senita), and *Stenocereus thurberi* (organ pipe, pitaya), also of the Sonoran Desert, to lower coverage. *Pereskia humboldtii*, a leafy, Andean cactus, was included as an outgroup (31).

Results and Discussion

The assembly of saguaro's 1.4-Gb genome (32) from short read libraries (Table S1) spanned 980 Mb, with a scaffold N50 of 61.5 kb (Table S2). Transcriptomes were also assembled (Table S3) and used with other evidence to annotate the saguaro genome. The saguaro genome contains 28,292 protein coding genes; 58% of the assembly consists of transposable elements and retroviral and repeated sequences (Tables S2 and S4). Assemblies from single libraries of the other four cacti were more fragmented (Table S5).

The high completeness of the gene space in the saguaro assembly (Table S2) let us construct 4,436 alignments of orthologous genes across the five cacti, each of which contained two or more exons for all five taxa, and contiguous introns. Based on gene tree confidence levels estimated from alignments containing all nucleotide positions, sets of alignments differing in phylogenetic robustness ("gene confidence sets") were compiled for downstream analyses. For example, the 90% gene confidence set comprised 458 genes, with gene trees having maximum likelihood bootstrap support value above 90% for all clades (Table 1). These gene sets provide a control for the effect of weakly supported gene trees on gene tree discordance estimates (33).

The most robust 90% gene confidence set implies levels of gene tree discordance between Pachycereeae genera (Fig. 2*A*) comparable to those seen in closely related short-lived angiosperm genera such as *Solanum* sect. *Lycopersicon* [time scale of 2 million years (MYR); ref. 24] and the *Oryza* AA genome clade (2.5 MYR: ref. 22). Species trees constructed using both gene tree-based (MP-EST; ref. 34) and alignment-based methods (BPP; ref. 35) were the same for all gene tree confidence sets and for three different partitioning schemes for the alignments by codon position and intron (*Materials and Methods* and Fig. 2*B*). An exhaustive search of tree space indicated a single optimal species tree under the MP-EST pseudolikelihood criterion, and three replicate runs of each MCMC chain in BPP from random starting trees generated this same species tree for all five gene confidence sets, indicating convergence. This optimal tree agreed with previous molecular phylogenetic analyses of mostly plastid genes (6, 28, 29). In the 80% and 90% gene confidence sets, 61% and 63% of gene trees, respectively, agreed with the

Table 1. Gene tree discordance relative to the Pachycereeae species tree (Fig. 2*B*)

Gene confidence set	90% set	80% set	70% set	60% set	50% set
N genes	458	786	1,065	1,668	2,291
Gene tree concordance, %*					
Edge A†	76	75	74	68	65
Edge B	77	75	75	73	70
Whole tree	63	61	58	52	47
MP-EST edge length‡					
Edge A	1.01	0.96	0.94	0.78	0.69
Edge B	1.22	1.21	1.15	1.04	0.93
BPP edge length§					
All partition					
Edge A	1.23	1.14	1.09	0.95	0.88
Edge B	1.22	1.22	1.18	1.14	1.10
Neutral partition					
Edge A	1.24	1.16	1.09	0.97	0.89
Edge B	1.28	1.29	1.24	1.21	1.18
Third partition					
Edge A	1.22	1.11	1.08	0.95	0.87
Edge B	1.32	1.34	1.24	1.18	1.12

*Agreement of gene tree edge bipartitions or entire gene tree with species tree.
†See Fig. 2*B* for location of edges.
‡In coalescent time units. 1 CTU = $2N_e$ generations.
§In CTUs estimated from $2D/\theta$, where D and θ are sequence divergence-scaled edge length and scaled effective population size, respectively, taken from BPP output.

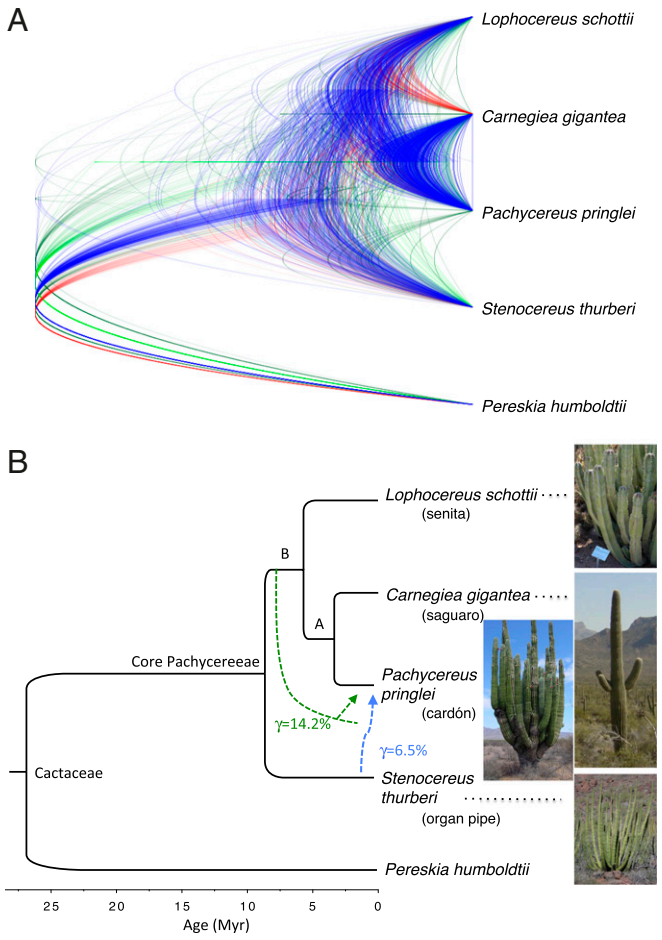

Fig. 2. (*A*) Topological discordance in the 90% gene confidence set of 458 gene trees visualized in DensiTree (78). Blue trees are the most frequent, followed by red and shades of green. (*B*) Species tree inferred with BPP and MP-EST, which was the same for all gene sets, with divergence times estimated using the Neutral partition. Blue dashed line indicates reticulation, having inferred inheritance probability, γ, on the optimal PhyloNet analysis of the same set of gene trees. Green dashed line indicates position of reticulation on the next best inferred network (Fig. S1). The origin of this reticulation edge earlier than its endpoint implies the existence of an extinct or unsampled taxon.

species tree (Table 1). Estimates of species tree edge lengths for these gene confidence sets in BPP ranged from 1.11 to 1.24 for edge A and from 1.22 to 1.34 for edge B, in "coalescent time units" (= $T/2N_e$, where T is measured in generations), depending on data partition, with slightly lower values estimated by MP-EST (Table 1).

Introgression may also contribute to gene tree discordance. Network reconstruction allowing both ILS and gene flow in PhyloNet (36) indicated substantial support for a network model with one reticulation vs. a tree model (ΔAIC = 17.0 and 28.6 for the 90% and 80% gene sets, respectively; ref. 37), implicating introgression into *P. pringlei* from either *S. thurberi* or some more closely related but unsampled or extinct taxon (Fig. 2*B* and Fig. S1). The best two networks were the same in both gene confidence sets (although the ranking reversed; Fig. S1). Moreover, the "major tree" within these networks (*Materials and Methods*) was the same as that inferred by species tree methods, and the inheritance probability estimate from the 90% gene confidence set, which indicates the level of introgression, was low (6.5% and 14.2% for the optimal and next best networks in the 90% gene confidence set; although higher in one of the two networks from the 80% gene set). The two optimal networks were consistently recovered in multiple searches from different starting networks, and in each gene set, the two results were substantially better than other suboptimal networks (Fig. S1).

Intergeneric hybrids are well known in cacti (1, 5), including rare hybrids between *P. pringlei* and *Bergerocactus emoryi* in Baja California in a narrow zone where the two species are sympatric (1, 38). *Bergerocactus* (not sampled here) is more closely related to *Carnegiea*, *Lophocereus*, and *Pachycereus* than to *Stenocereus* in the most complete cactus phylogeny to date (6). *P. pringlei* is possibly a recent autotetraploid (39), so postzygotic inviability and sterility barriers (40) would have had to have been overcome if the introgression took place following tetraploidy.

Owing to the limited level of introgression, we inferred demographic history using BPP assuming that ILS is the primary cause of discordance. The neutral mutation rate was estimated from the substitution rate in the "neutral" partition of the alignments (Table S6) assuming a root age of the tree at 26.88 MYR (*Materials and Methods*). Estimates ranged from $\mu_G = 5.98$ to 6.07×10^{-8} per site for each generation across the five gene confidence sets. An alternative estimate based on pairwise synonymous K_s distances in CDS regions between *Carnegiea* and *Pereskia* from all 4,426 alignments was somewhat higher at $\mu_G = 8.75 \times 10^{-8}$ per site per generation (Table S7). Per year rates are comparable to several angiosperm tree species (41–43), but generation times in the cacti, ranging from 20 to 75 y (44–47), are 2–12 times longer. The BPP estimates of scaled mutation rate ($\theta = 4\ N_e\mu_G$), imply an N_e of 24,000–39,000 and 31,000–36,000 for edges A and B, respectively, over the 15 gene sets and partitions (Table S6). Combining the BPP estimates of θ with the pairwise K_s estimate of mutation rate produces ancestral N_e of ~2/3 the BPP estimates. These ancestral N_e values are higher than estimates of recent N_e for some perennial angiosperms [*Populus trichocarpa*: ~4,000–6,000 (48); *Eucalyptus grandis*: ~11,000 (41); *Amborella trichopoda*: 5,000 (49)] but comparable to *Populus balsamifera* (44,000–59,000; ref. 50); and much less than *Pinus taeda* (560,000; ref. 51).

The probability that a rooted gene tree with three taxa disagrees with the species tree that contains it is $2/3\ e^{-T/2Ne}$, where T is the time in generations along the internal edge of the tree (52). This can be high if T is small, either because divergence time in years is small or generation time is large. Thus, the level of gene tree discordance we inferred in columnar cacti having N_e of 25–40,000, along edges with a duration of 2–3 million years would be far too high were it not for the long generation times of these plants, similar to findings in long-lived conifers (53).

Gene tree discordance has a strong impact on the distribution of protein sequence variation among these genera of cacti in generating hemiplasy. Taking the 80% and 90% gene confidence sets as samples of potentially phenotypically relevant amino acid sequence variation across these genomes, we partitioned the homoplastic amino acid sites in genes on the species tree into three parts: the fraction arising from homoplasy on concordant gene trees (Fig. 1, *Upper Right*), the fraction arising solely from gene trees with less homoplasy than the species tree but that are discordant with the species tree (hemiplasy: Fig. 1, *Lower Left*), and the small fraction that arises from homoplasy on discordant gene trees that is equally homoplastic on the species tree (which as yet has no term defined: Fig. 1, *Lower Right*). Hemiplasy accounts for 58–63% of all apparent homoplasy (Table 2).

As examples, consider two genes from the 90% gene confidence set having strongly discordant gene trees. Both play potential roles in the physiological adaptation of cacti to arid environments. The saguaro nuclear gene annotated as a chloroplastic NADP-dependent malate dehydrogenase (MDH: Cgig1_18427) is the only 1 of 11 MDH isoforms in the saguaro annotation that is NADP-dependent. It catalyzes the reduction of oxaloacetate to malate in the chloroplast (54) and appears to participate in the fixation of carbon dioxide under both light and dark conditions

EVOLUTION

Table 2. Homoplasy and hemiplasy in amino acid alignments of genes

Gene confidence set	90% set	80% set
Concordant gene trees		
Variable sites	11,641	17,349
Informative sites	1,091	1,549
Homoplastic sites*	154 (27.8%)†	234 (31.0%)†
Discordant gene trees		
Variable sites	6,810	11,572
Informative sites	703	1,102
Homoplastic sites–not hemiplastic‡	79 (14.2%)†	46 (6.1%)†
Homoplastic sites–hemiplastic§	320 (57.9%)†	474 (62.9%)†
Total homoplastic sites on species tree	553	754

*See Fig. 1, *Upper Right*.
†% = percent of all homoplastic sites.
‡Fig. 1, *Lower Right*.
§Fig. 1, *Lower Left*.

(55). Together with the cytosolic NAD-dependent MDH, these two isoforms may be primarily responsible for malate formation during dark CO_2 fixation in crassulacean acid metabolism (CAM) plants (56). The MDH gene tree has a single replacement of an ancestral serine with an alanine, but the gene tree is discordant with the species tree, making the alanines appear to have evolved twice on the species tree, when they are in fact hemiplastic.

The gene annotated as a DNAJ JJJ1 homolog (Cgig1_00352) contains a DNAJ domain involved in heat shock protein interactions. Columnar cacti must be able to tolerate internal tissue temperatures that can exceed 50 °C in the summer (5). The gene tree is discordant with the species tree, and 19 of its 20 potentially informative amino acid replacements exhibit homoplasy on the species tree, but only one does on the gene tree, meaning 18 sites are hemiplastic. A search of the National Center for Biotechnology Information (NCBI) CDD database (57) shows that two of these hemiplastic sites are in the conserved DNAJ domain proper, although neither involves known HSP70 interaction sites.

Hemiplasy is not restricted to the coding sequence in this gene. The phylogeny of the 1,000 bp upstream of the start codon, which often contains proximal conserved regulatory elements in plants (58), is the same discordant tree as is found for the coding region. Of 22 potentially informative nucleotide sites there, 21 are homoplastic on the species tree, whereas only two are on the gene tree. Most homoplasy in this noncoding region on the species tree is thus also due to hemiplasy. All of these cases suggest caution in viewing multiple origins of the same amino acid or nucleotide in a species tree as possible evidence of convergent evolution (18).

These conclusions depend on the robustness of various model assumptions. The extent of gene tree discordance and hemiplasy was estimated from gene tree topologies inferred with an HKY substitution model in PAUP*. Estimates of tree topology are generally more robust than estimates of rates, especially at low sequence divergences (59–61), where simple models sometimes even outperform more general ones (62). Mean pairwise nonsynonymous and synonymous distances between the most divergent Pachycereeae were only 1.3% and 4.1% respectively, and about one-half of the 4,436 sequence alignments had bootstrap values below 50%, a consequence of low sequence divergence and short length. Substitution model therefore likely had little impact on estimates of the prevalence of discordance and hemiplasy.

However, our explanation for these levels of hemiplasy in terms of generation time rests on estimates of divergence times and N_e obtained from BPP, which adds the assumptions of panmixia within and no gene flow between species, constant population sizes within tree edges, and no linkage (61). To examine the impact of BPP's Jukes–Cantor clock model on its estimates of divergence times and effective population sizes, we evaluated more general models with subsets of the data using BEAST (63), finding that the parameter effect sizes were small (*SI Materials and Methods* and Table S8). Other assumptions were more difficult to test. Panmixia will not be strictly true even for outcrossing columnar cacti, but gene flow is likely high within three of the four Pachycereeae that are bat pollinated (e.g., *S. thurberi*; ref. 64). We did find limited gene flow between species in PhyloNet analyses, and this can lead to overestimates of population sizes and underestimates of divergence times when using the multispecies coalescent approach (65), but because the rates of ILS correlate with the product of population size and time, these biases potentially counteract each other. However, it is unlikely that population sizes have remained constant within species tree edges. Paleoclimatic and packrat midden evidence document the ebb and flow of Sonoran Desert plant communities during Pleistocene glacial and interglacial periods (66). BPP's divergence time estimates are biased toward the recent in "more extreme bottleneck scenarios" of population history (67), but shorter edge durations in Pachycereeae would tend to make the observed level of ILS even higher than estimated. Finally, linkage is unlikely to be problematic. Our 80% gene confidence set has an average of 1.8 Mb between genes. Linkage disequilibrium in long-lived, outcrossing angiosperms typically decays in distances from a few kb to 50 kb (41, 48).

The phylogenomic history of saguaro and its relatives exhibits extensive gene tree discordance due to ILS and low rates of introgression. Comparable findings have been seen in rapid and/or very recent radiations (20, 22, 24), but in Pachycereeae, ILS acts at long time scales, between taxonomically divergent genera with very long generation times. A consequence of this discordance is elevated levels of apparent homoplasy in the species tree. The connection between apparent genomic homoplasy arising from ILS and apparent phenotypic homoplasy is probably strongest for traits with a simple genetic architecture (18), such as those involving the function or regulation of a single enzyme (21, 68). In plants, enzymes in biosynthetic pathways for floral pigments or defensive compounds are good candidates (69, 70). Notably, the taxonomic distribution of alkaloids, triterpenes, and sterols have played a role in the systematics of Pachycereeae (30). When hemiplasy arises because of introgression, genetic architecture may be less of an issue because multiple loci may undergo gene flow almost simultaneously (18). In plants, even complex traits with multiple components and fitness effects have been adaptively introgressed (71). Collectively, this body of evidence lends plausibility to the hypothesis that the phenotypic effects of genomic hemiplasy may have exacerbated the long-standing problem of inferring relationships in these charismatic cacti.

Materials and Methods

Sampling, Genome Sequencing, and Annotation. For details of taxa sampled, library construction, genome sequencing, assembly, and annotation, see *SI Materials and Methods*.

Phylogenomic Analyses.

Gene alignments and gene trees. Sets of gene trees were inferred from gene alignments constructed from the genome assemblies with custom PERL scripts. CDS and intron regions were extracted from the saguaro genome based on its annotation. Each region was used as a query in blastn searches (72) against the other four cactus assemblies, with an e value cutoff of 10^{-10} for the intron searches. To enrich for orthologs, (*i*) a CDS region was kept for further processing if it returned exactly one hit for all four taxa and (*ii*) the gene was kept if and only if at least two CDS regions from the same gene passed test (*i*).

We then used predicted saguaro amino acid sequences within the retained CDS regions in pairwise tblastn runs against the nucleotide hits found in each of the four cacti. Thereby, we obtained subject sequences in the same frame as the saguaro query. We used Muscle (73) to align these protein regions and then tranalign (74) to align the underlying nucleotide sequences. Each gene thus consists of CDS regions that have sequence data present for all taxa. Given at least one such region, introns were then evaluated. An intron region was kept if it returned exactly four hits, one per taxon, and each hit covered at least 50%

of the query saguaro sequence. Any alignment in which at least one taxon had more than 50% gaps or missing data was excluded from further consideration.

Three partitions of each gene nucleotide alignment were prepared: one having just third positions in codons ("Third"), one having third positions plus introns ("Neutral"); and one having all nucleotide positions ("All"). Bootstrap maximum likelihood majority rule trees for the All positions alignment were constructed with PAUP* v. 4.0a with an HKY85 model (75). "Gene confidence sets" were compiled based on the quality of their trees (e.g., the "90%" set is all genes for which the minimum bootstrap value is 90% for all clades). Sets were assembled at 50, 60, 70, 80, and 90% levels and rooted with *Pereskia*.

Species tree inference. Species trees were constructed first by an alignment-based Bayesian method, BPP v.3.3a (35, 61), in which the input was the alignments from a gene confidence set and partition. Convergence of Markov chains was checked by running three chains from random starting trees for each of the five gene confidence sets (burnin = 4,000; number of generations = 40,000). This is aimed at avoiding multiple optima in the posterior (76) and may be more informative than examining acceptance rates or effective sample sizes (35).

The prior for the scaled root divergence time, τ_{root}, was set to a gamma distribution, based on the mean and variance of pairwise K_s distances between saguaro and *Pereskia* (Table S7). The prior for scaled population size, θ_{root}, was set to a gamma distribution with a mean estimated from the modern nucleotide diversity of saguaro, which was estimated, using the program PSMC (77) applied to the saguaro genome scaffolds ≥100 kb in length, to be $\pi = 0.00146$ (variance set to half the mean). The rate variation prior was set to a Dirichlet prior with parameter $\alpha = 2$.

A second, tree-based pseudolikelihood method, MP-EST v. 1.5 (34), was used with an input of all rooted gene trees constructed from the All partition from each gene confidence set. In each, five replicate searches from random starting species trees were done. To check for multiple optima, an exhaustive enumeration of pseudolikelihood scores for all possible ingroup rooted species trees was done by supplying MP-EST with those 15 trees.

Gene tree discordance. Discordance between gene trees and the inferred species tree was assayed on a clade basis and for the entire gene tree jointly using PAUP*. Discordant gene trees were binned into groups of identical topology by comparing them to all 14 possible rooted discordant binary trees for five taxa using a custom PERL script. Topological discordance in the All partition was visualized using DensiTree (78) by making the gene trees in the 90% gene confidence set ultrametric using a clock model in PAUP* and scaling root ages to 1.0 (Fig. 2*A*).

Introgression and network reconstruction. We used InferNetwork_ML in PhyloNet (36) to reconstruct optimal phylogenetic networks using maximum likelihood, using as input gene trees from the All partition [invoked with "InferNetwork_ML (all) *h* -n 10 -di -o -po -x 20 -s starting_tree", where *h* is the maximum allowed number of reticulation events]. Values of *h* of 0 (a tree) and 1 were each run from three different starting topologies. Each inferred reticulation node has two incident edges with inheritance probabilities γ and $1 - \gamma$. The tree imbedded in the network, obtained by following the path along the larger of the two inheritance probabilities, is called the major tree.

To compare solutions, we used the AIC information criteria (37): AIC = $2k - 2 \log L$, where k is the number of parameters in the model and L is the maximum likelihood value of the model. To assess the statistical support for the optimal network, we used PhyloNet's parametric bootstrap procedure with the same search parameters as used previously.

Demographic inference. We used BPP to infer divergence times and effective population sizes conditional on the species tree found above. To ensure convergence of the Markov chains, we increased the number of generations to 400,000 (burnin = 40,000), leading to all parameter estimates having an acceptable effective sample size (ESS) >500 (79) and acceptance rates lying in the range 0.25–0.40. We also examined parameter traces in Tracer (63). The raw parameter output of BPP is in units of sequence divergence, $D = T\mu$, where T is the time in generations and μ is the per generation mutation rate per site, and scaled effective population size, $\theta = 4N_e\mu$. Note that $2D/\theta = T/2N_e$ is the edge length in "coalescent time units."

Quantifying Hemiplasy. For any site in an alignment, let m_S and m_G be the inferred number of state changes on the species and gene trees respectively. If the two trees are concordant, $m_S = m_G$, but if the trees are discordant, then the number of changes may differ. The simplest case of hemiplasy is $m_G = 1$ but $m_S > 1$: A homology on the gene tree is a homoplasy on the species tree (Fig. 1, *Lower Left*). A site may also exhibit homoplasy on the gene tree: $m_G > 1$ (Fig. 1, *Lower Right*). We define a hemiplastic site as one in which $m_S > m_G$. The exceptional case of $m_G > 1$ but $m_S = m_G$ is "homoplasy but not hemiplasy."

Ancestral states for each residue in each protein sequence alignment in the most robust 80% and 90% gene confidence sets were reconstructed using parsimony with PAUP*, and m_S and m_G were computed for all potentially parsimony informative sites. Sites homoplastic on the species tree for concordant gene trees, and sites homoplastic or hemiplastic for nonconcordant gene trees, were tallied by a custom PERL script.

Neutral Mutation Rate Estimates. We estimated neutral mutation rate from BPP output as sequence divergence from root to tip divided by the crown group age of cacti, since the assumed model is ultrametric. We also used codeml (80) to infer pairwise synonymous divergences in the coding regions across the entire set of gene trees, using the crown age of Cactaceae estimated at 26.88 MYR (81).

ACKNOWLEDGMENTS. We thank S. Kumar, T. Hernández-Hernández, B. Rannala, K. Steele, F. Tax, L. Venable, and D. Zwickl for discussion, and the Tumamoc Hill Reserve, Tucson, and the Desert Botanical Garden, Phoenix, for permission to collect material. Funding was provided by the University of Arizona–Universidad Nacional Autónoma de México Consortium for Drylands Research, the Tucson Cactus and Succulent Society, and Arizona State University's College of Liberal Arts and Sciences and the School of Life Sciences. A.B. received sabbatical support at the University of Arizona from DGAPA-Universidad Nacional Autónoma de México and by Programa de Apoyo a Proyectos de Investigación e Innovación Tecnológica, Universidad Nacional Autónoma de México Grant IN213814. N.K.W. received support from NIH Grant R35GM119816.

1. Anderson EF (2001) *The Cactus Family* (Timber, Portland, OR), p 776.
2. Hunt D (2006) *The New Cactus Lexicon* (Remous, Milborne Port, UK).
3. Schumann K (1903) *Gesamtbeschreibung der Kakteen (Monographia Cactacearum)* (J. Neumann, Neudamm, Germany), 2nd Ed, p 832.
4. Backeberg C (1966) *Das Kakteenlexicon* (Gustav Fischer, Stuttgart).
5. Gibson AC, Nobel PS (1986) *The Cactus Primer* (Harvard Univ Press, Cambridge, MA), p 286.
6. Hernández-Hernández T, et al. (2011) Phylogenetic relationships and evolution of growth form in Cactaceae (Caryophyllales, Eudicotyledoneae). *Am J Bot* 98:44–61.
7. Sanderson M, Hufford L, eds (1996) *Homoplasy: The Recurrence of Similarity in Evolution* (Academic, New York).
8. Britton NL, Rose JN (1920) *The Cactaceae* (Dover, Mineola, NY), 2nd Ed.
9. Nyffeler R (2002) Phylogenetic relationships in the cactus family (Cactaceae) based on evidence from *trnK/ matK* and *trnL-trnF* sequences. *Am J Bot* 89:312–326.
10. Wake DB (1991) Homoplasy–The result of natural-selection, or evidence of design limitations. *Am Nat* 138:543–567.
11. Cota JH, Wallace RS (1997) Chloroplast DNA evidence for divergence in *Ferocactus* and its relationships to North American columnar cacti (Cactaceae: Cactoideae). *Syst Bot* 22: 529–542.
12. Griffith MP, Porter JM (2009) Phylogeny of Opuntioideae (Cactaceae). *Int J Plant Sci* 170:107–116.
13. Nyffeler R, Eggli U (2010) A farewell to dated ideas and concepts: Molecular phylogenetics and a revised suprageneric classification of the family Cactaceae. *Schumannia* 6:109–149.
14. Porter JM, Kinney M, Heil KD (2000) Relationships between *Sclerocactus* and *Toumeya* (Cactaceae) based in chloroplast *trnL-F* sequences. *Haseltonia* 7:8–23.
15. Bárcenas RT, Yesson C, Hawkins JA (2011) Molecular systematics of the Cactaceae. *Cladistics* 27:470–489.
16. Goettsch B, et al. (2015) High proportion of cactus species threatened with extinction. *Nat Plants* 1:15142.
17. Ogburn RM, Edwards EJ (2009) Anatomical variation in Cactaceae and relatives: Trait lability and evolutionary innovation. *Am J Bot* 96:391–408.
18. Hahn MW, Nakhleh L (2016) Irrational exuberance for resolved species trees. *Evolution* 70:7–17.
19. Avise JC, Robinson TJ (2008) Hemiplasy: A new term in the lexicon of phylogenetics. *Syst Biol* 57:503–507.
20. Suh A, Smeds L, Ellegren H (2015) The dynamics of incomplete lineage sorting across the ancient adaptive radiation of neoavian birds. *PLoS Biol* 13:e1002224.
21. Storz JF (2016) Causes of molecular convergence and parallelism in protein evolution. *Nat Rev Genet* 17:239–250.
22. Zwickl DJ, Stein JC, Wing RA, Ware D, Sanderson MJ (2014) Disentangling methodological and biological sources of gene tree discordance on *Oryza* (Poaceae) chromosome 3. *Syst Biol* 63:645–659.
23. Fontaine MC, et al. (2015) Mosquito genomics. Extensive introgression in a malaria vector species complex revealed by phylogenomics. *Science* 347:1258524.
24. Pease JB, Haak DC, Hahn MW, Moyle LC (2016) Phylogenomics reveals three sources of adaptive variation during a rapid radiation. *PLoS Biol* 14:e1002379.
25. Yu Y, Dong J, Liu KJ, Nakhleh L (2014) Maximum likelihood inference of reticulate evolutionary histories. *Proc Natl Acad Sci USA* 111:16448–16453.
26. Rasmussen MD, Kellis M (2012) Unified modeling of gene duplication, loss, and coalescence using a locus tree. *Genome Res* 22:755–765.
27. Gibson AC, Spencer KC, Bajaj R, McLaughlin JL (1986) The ever-changing landscape of cactus systematics. *Ann Mo Bot Gard* 73:532–555.
28. Hartmann S, Nason JD, Bhattacharya D (2002) Phylogenetic origins of *Lophocereus* (Cactaceae) and the senita cactus-senita moth pollination mutualism. *Am J Bot* 89:1085–1092.
29. Arias S, Terrazas T, Cameron K (2003) Phylogenetic analysis of *Pachycereus* (Cactaceae, Pachycereeae) based on chloroplast and nuclear DNA sequences. *Syst Bot* 28:547–557.
30. Gibson AC, Horak KE (1978) Systematic anatomy and phylogeny of Mexican columnar cacti. *Ann Mo Bot Gard* 65:999–1057.

31. Edwards EJ, Nyffeler R, Donoghue MJ (2005) Basal cactus phylogeny: Implications of *Pereskia* (Cactaceae) paraphyly for the transition to the cactus life form. *Am J Bot* 92: 1177–1188.
32. Bennett MD, Leitch IJ (2012) Plant DNA C-values database (release 6.0, December 2012). Available at data.kew.org/cvalues/. Accessed December 1, 2016.
33. Xu B, Yang Z (2016) Challenges in species tree estimation under the multispecies coalescent model. *Genetics* 204:1353–1368.
34. Liu L, Yu L, Edwards SV (2010) A maximum pseudo-likelihood approach for estimating species trees under the coalescent model. *BMC Evol Biol* 10:302.
35. Rannala B, Yang Z (2017) Efficient Bayesian species tree inference under the multispecies coalescent. *Syst Biol* 66:823–842.
36. Than C, Ruths D, Nakhleh L (2008) PhyloNet: A software package for analyzing and reconstructing reticulate evolutionary relationships. *BMC Bioinformatics* 9:322.
37. Burnham KP, Anderson DR (2010) *Model Selection and Multi-Model Inference* (Springer, New York), 3rd Ed, p 496.
38. Moran R (1962) *Pachycereus orcuttii*—A puzzle solved. *Cactus Succulent J (US)* 34:88–94.
39. Murawski DA, Fleming TH, Ritland K, Hamrick JL (1994) Mating system of *Pachycereus pringlei*: An autotetraploid cactus. *Heredity* 72:86–94.
40. Husband BC, Sabara HA (2004) Reproductive isolation between autotetraploids and their diploid progenitors in fireweed, *Chamerion angustifolium* (Onagraceae). *New Phytol* 161:703–713.
41. Silva-Junior OB, Grattapaglia D (2015) Genome-wide patterns of recombination, linkage disequilibrium and nucleotide diversity from pooled resequencing and single nucleotide polymorphism genotyping unlock the evolutionary history of *Eucalyptus grandis*. *New Phytol* 208:830–845.
42. Sollars ESA, et al. (2017) Genome sequence and genetic diversity of European ash trees. *Nature* 541:212–216.
43. Luo MC, et al. (2015) Synteny analysis in Rosids with a walnut physical map reveals slow genome evolution in long-lived woody perennials. *BMC Genomics* 16:707.
44. Steenbergh W, Lowe C (1983) *Ecology of the Saguaro: Growth and Demography, Part 3*, National Park Service Scientific Monograph Series (Government Printing Office, Washington, DC).
45. Parker KC (1988) Growth-rates of *Stenocereus thurberi* and *Lophocereus schottii* in Southern Arizona. *Bot Gaz* 149:335–346.
46. Parker KC (1989) Height structure and reproductive characteristics of senita, *Lophocereus schottii* (Cactaceae), in Southern Arizona. *Southwest Nat* 34:392–401.
47. Dimmitt M (2016) Cactaceae (the cactus family). Available at www.desertmuseum.org/books/nhsd_cactus_.php. Accessed January 1, 2017.
48. Slavov GT, et al. (2012) Genome resequencing reveals multiscale geographic structure and extensive linkage disequilibrium in the forest tree *Populus trichocarpa*. *New Phytol* 196:713–725.
49. Albert VA, et al.; Amborella Genome Project (2013) The *Amborella* genome and the evolution of flowering plants. *Science* 342:1241089.
50. Olson MS, et al. (2010) Nucleotide diversity and linkage disequilibrium in balsam poplar (*Populus balsamifera*). *New Phytol* 186:526–536.
51. Brown GR, Gill GP, Kuntz RJ, Langley CH, Neale DB (2004) Nucleotide diversity and linkage disequilibrium in loblolly pine. *Proc Natl Acad Sci USA* 101:15255–15260.
52. Hudson RR (1983) Testing the constant-rate neutral allele model with protein sequence data. *Evolution* 37:203–217.
53. Zhou Y, et al. (2017) Importance of incomplete lineage sorting and introgression in the origin of shared genetic variation between two closely related pines with overlapping distributions. *Heredity (Edinb)* 118:211–220.
54. Scheibe R (1987) NADP+-malate dehydrogenase in C3-plants: Regulation and role of a light-activated enzyme. *Physiol Plant* 71:393–400.
55. Cushman JC (1993) Molecular cloning and expression of chloroplast NADP-malate dehydrogenase during Crassulacean acid metabolism induction by salt stress. *Photosynth Res* 35:15–27.
56. Mallona I, Egea-Cortines M, Weiss J (2011) Conserved and divergent rhythms of crassulacean acid metabolism-related and core clock gene expression in the cactus *Opuntia ficus-indica*. *Plant Physiol* 156:1978–1989.
57. Marchler-Bauer A, et al. (2015) CDD: NCBI's conserved domain database. *Nucleic Acids Res* 43:D222–D226.
58. Haudry A, et al. (2013) An atlas of over 90,000 conserved noncoding sequences provides insight into crucifer regulatory regions. *Nat Genet* 45:891–898.
59. Zharkikh A (1994) Estimation of evolutionary distances between nucleotide sequences. *J Mol Evol* 39:315–329.
60. Bos DH, Posada D (2005) Using models of nucleotide evolution to build phylogenetic trees. *Dev Comp Immunol* 29:211–227.
61. Yang ZH (2015) The BPP program for species tree estimation and species delimitation. *Curr Zool* 61:854–865.
62. Doerr D, Gronau I, Moran S, Yavneh I (2012) Stochastic errors vs. modeling errors in distance based phylogenetic reconstructions. *Algorithms Mol Biol* 7:22.
63. Bouckaert R, et al. (2014) BEAST 2: A software platform for Bayesian evolutionary analysis. *PLOS Comput Biol* 10:e1003537.
64. Bustamante E, Búrquez A, Scheinvar E, Eguiarte LE (2016) Population genetic structure of a widespread bat-pollinated columnar cactus. *PLoS One* 11:e0152329.
65. Leaché AD, Harris RB, Rannala B, Yang Z (2014) The influence of gene flow on species tree estimation: A simulation study. *Syst Biol* 63:17–30.
66. McAuliffe JR, Van Devender TR (1998) A 22,000-year record of vegetation change in the north-central Sonoran Desert. *Palaeogeogr Palaeoclimatol Palaeoecol* 141:253–275.
67. Barley AJ, Brown JM, Thomson RC (2017) Impact of model violations on the inference of species boundaries under the multispecies coalescent. *Syst Biol*, 10.1093/sysbio/syx073.
68. Lang M, et al. (2012) Mutations in the neverland gene turned *Drosophila pachea* into an obligate specialist species. *Science* 337:1658–1661.
69. Huang R, O'Donnell AJ, Barboline JJ, Barkman TJ (2016) Convergent evolution of caffeine in plants by co-option of exapted ancestral enzymes. *Proc Natl Acad Sci USA* 113:10613–10618.
70. Brockington SF, et al. (2015) Lineage-specific gene radiations underlie the evolution of novel betalain pigmentation in Caryophyllales. *New Phytol* 207:1170–1180.
71. Whitney KD, Randell RA, Rieseberg LH (2010) Adaptive introgression of abiotic tolerance traits in the sunflower *Helianthus annuus*. *New Phytol* 187:230–239.
72. Altschul SF, et al. (1997) Gapped BLAST and PSI-BLAST: A new generation of protein database search programs. *Nucleic Acids Res* 25:3389–3402.
73. Edgar RC (2004) MUSCLE: A multiple sequence alignment method with reduced time and space complexity. *BMC Bioinformatics* 5:113.
74. Rice P, Longden I, Bleasby A (2000) EMBOSS: The European molecular biology open software suite. *Trends Genet* 16:276–277.
75. Swofford DL (2002) PAUP*. Phylogenetic Analysis Using Parsimony (*and Other Methods) (Sinauer, Sunderland, MA), Version 4.0.
76. Whidden C, Matsen FA, 4th (2015) Quantifying MCMC exploration of phylogenetic tree space. *Syst Biol* 64:472–491.
77. Li H, Durbin R (2011) Inference of human population history from individual whole-genome sequences. *Nature* 475:493–496.
78. Bouckaert RR (2010) DensiTree: Making sense of sets of phylogenetic trees. *Bioinformatics* 26:1372–1373.
79. Drummond A, Bouckaert RR (2015) *Bayesian Evolutionary Analysis with BEAST* (Cambridge Univ Press, Cambridge, UK), p 260.
80. Yang Z (2007) PAML 4: Phylogenetic analysis by maximum likelihood. *Mol Biol Evol* 24:1586–1591.
81. Hernández-Hernández T, Brown JW, Schlumpberger BO, Eguiarte LE, Magallón S (2014) Beyond aridification: Multiple explanations for the elevated diversification of cacti in the New World Succulent Biome. *New Phytol* 202:1382–1397.
82. Sanderson MJ, et al. (2015) Exceptional reduction of the plastid genome of saguaro cactus (*Carnegiea gigantea*): Loss of the ndh gene suite and inverted repeat. *Am J Bot* 102:1115–1127.
83. Joshi N, Fass J (2011) Sickle: A Sliding-Window, Adaptive, Quality-Based Trimming Tool for FastQ Files, Version 1.33. Available at https://github.com/najoshi/sickle. Accessed December 1, 2016.
84. Luo R, et al. (2012) SOAPdenovo2: An empirically improved memory-efficient short-read de novo assembler. *Gigascience* 1:18.
85. Heo Y, Wu XL, Chen D, Ma J, Hwu WM (2014) BLESS: Bloom filter-based error correction solution for high-throughput sequencing reads. *Bioinformatics* 30:1354–1362.
86. Schmieder R, Edwards R (2011) Quality control and preprocessing of metagenomic datasets. *Bioinformatics* 27:863–864.
87. Chapman JA, et al. (2011) Meraculous: De novo genome assembly with short paired-end reads. *PLoS One* 6:e23501.
88. Kajitani R, et al. (2014) Efficient de novo assembly of highly heterozygous genomes from whole-genome shotgun short reads. *Genome Res* 24:1384–1395.
89. Sahlin K, Vezzi F, Nystedt B, Lundeberg J, Arvestad L (2014) BESST–Efficient scaffolding of large fragmented assemblies. *BMC Bioinformatics* 15:281.
90. Hunt M, et al. (2013) REAPR: A universal tool for genome assembly evaluation. *Genome Biol* 14:R47.
91. Parra G, Bradnam K, Ning Z, Keane T, Korf I (2009) Assessing the gene space in draft genomes. *Nucleic Acids Res* 37:289–297.
92. Simão FA, Waterhouse RM, Ioannidis P, Kriventseva EV, Zdobnov EM (2015) BUSCO: Assessing genome assembly and annotation completeness with single-copy orthologs. *Bioinformatics* 31:3210–3212.
93. Novák P, Neumann P, Pech J, Steinhaisl J, Macas J (2013) RepeatExplorer: A Galaxy-based web server for genome-wide characterization of eukaryotic repetitive elements from next-generation sequence reads. *Bioinformatics* 29:792–793.
94. Copetti D, et al. (2015) RiTE database: A resource database for genus-wide rice genomics and evolutionary biology. *BMC Genomics* 16:538.
95. Flutre T, Duprat E, Feuillet C, Quesneville H (2011) Considering transposable element diversification in de novo annotation approaches. *PLoS One* 6:e16526.
96. Nawrocki EP, Eddy SR (2013) Infernal 1.1: 100-fold faster RNA homology searches. *Bioinformatics* 29:2933–2935.
97. Schattner P, Brooks AN, Lowe TM (2005) The tRNAscan-SE, snoscan and snoGPS web servers for the detection of tRNAs and snoRNAs. *Nucleic Acids Res* 33:W686–W689.
98. Campbell MS, et al. (2014) MAKER-P: A tool kit for the rapid creation, management, and quality control of plant genome annotations. *Plant Physiol* 164:513–524.
99. Grabherr MG, et al. (2011) Full-length transcriptome assembly from RNA-Seq data without a reference genome. *Nat Biotechnol* 29:644–652.
100. Edgar RC (2010) Search and clustering orders of magnitude faster than BLAST. *Bioinformatics* 26:2460–2461.
101. Berardini TZ, et al. (2015) The Arabidopsis information resource: Making and mining the "gold standard" annotated reference plant genome. *Genesis* 53:474–485.
102. Apweiler R, et al. (2014) Activities at the Universal Protein Resource (UniProt). *Nucleic Acids Res* 42:D191–D198.
103. Korf I (2004) Gene finding in novel genomes. *BMC Bioinformatics* 5:59.
104. Eilbeck K, Moore B, Holt C, Yandell M (2009) Quantitative measures for the management and comparison of annotated genomes. *BMC Bioinformatics* 10:67.
105. Stanke M, Waack S (2003) Gene prediction with a hidden Markov model and a new intron submodel. *Bioinformatics* 19:ii215–ii225.
106. Finn RD, Clements J, Eddy SR (2011) HMMER web server: Interactive sequence similarity searching. *Nucleic Acids Res* 39:W29–W37.
107. Finn RD, et al. (2014) Pfam: The protein families database. *Nucleic Acids Res* 42:D222–D230.
108. Patro R, Duggal G, Love MI, Irizarry RA, Kingsford C (2017) Salmon provides fast and bias-aware quantification of transcript expression. *Nat Methods* 14:417–419.
109. Felsenstein J (1981) Evolutionary trees from DNA sequences: A maximum likelihood approach. *J Mol Evol* 17:368–376.