Reconciliation of evidence for the Portulacineae "backbone" phylogeny

MARK A. HERSHKOVITZ
Santiago, Chile
cistanthe@gmail.com

ABSTRACT

Portulacineae comprise a clade of eight ostensibly monophyletic families, four of which (Anacampserotaceae, Montiaceae, Portulacaceae s. str., and Talinaceae) and part of a fifth (Didiereaceae) had been classified traditionally in Portulacaceae s. lato. The clade also includes Basellaceae, Cactaceae, and Halophytaceae. While available evidence strongly supports recognition of major clades within Portulacineae, current analyses disagree with respect to relations among them, such that the Portulacineae "backbone" phylogeny remains "unresolved." The disagreements might owe in part to incongruent data and/or poor analysis and/or known theoretical shortcomings of the analytical methods. But I argue here that it reflects mostly the failure to appreciate the fundamental property of *living* organisms, viz. their inherent *determinism* consequent to autopoiesis. This property renders the evolutionary process as idiosyncratic. This, in turn, renders phylogeny *inherently* unpredictable and, strictly speaking, unrecoverable. I also emphasize that the hierarchical organization of organisms predicts that phylogeny should not be strictly tree-like. Nonetheless, evolutionary history is materially tangible, hence is within the realm of scientific inquiry. I make two proposals here. One is that (often futile) efforts to resolve phylogeny as a tree reflect a constitutive cognitive proclivity to resolve trees even when phylogeny is not tree-like and/or otherwise "resolvable." To mitigate this tendency, I propose that the objective of phylogenetic study should be *reconciliation* rather than resolution. In this way, the lack of tree-like phylogenetic resolution becomes useful knowledge. In this theoretical framework, I evaluate what can be considered tentatively *known* about the Portulacineae backbone phylogeny.
Key words: Portulacineae, phylogenetics, phylogenomics, hierarchy theory, autopoiesis, evolutionary idiosyncraticity, cognition, reconciliation

CITATION: Hershkovitz, M.[A.] 2021. Reconciliation of evidence for the Portulacineae "backbone" phylogeny. EcoevoRxiv. https://doi.org/10.32942/osf.io/yrdsf [this is version 11; check for latest version]

I. Introduction

In a companion to the present paper, I review evidence for the historical and current taxonomic distribution of perianth fate traits among Portulacineae. In current classifications, this suborder comprises eight evidently monophyletic families (Anacampserotaceae, Basellaceae, Cactaceae, Didiereaceae, Halophytaceae, Montiaceae [including Hectorellaceae], Portulacaceae, and Talinaceae; Hernández-Ledesma et al., 2015). In 19–20th Century taxonomies (e.g., McNeill, 1974; Cronquist, 1981; Carolin, 1987, 1993), current Anacampserotaceae, Montiaceae, Portulacaceae [s. str.], Talinaceae, and three current genera of Didiereaceae were classified in Portulacaceae [s. l.].

This present work emerges as the latest of preliminary papers spun off successively in the course of preparing what should have been an utterly simple and unexciting review of the taxonomy of the annual species of *Calandrinia* Kunth (sensu Hershkovitz, 1993, 2019a). The other spinoffs now

published,[1] totaling 150 pages, dealt with problems pertaining to the taxonomy and nomenclature of particular species as well as a persistent, yet inane, controversy regarding the circumscription of *Calandrinia* (Hershkovitz, 2020a–d; 2021a; cf. 2019a).

The present work extends to all of Portulacineae. It was motivated in part by the observation that perianth fate figured significantly in the generic taxonomy of classically circumscribed Portulacaceae. In particular, the 19–20th Century polyphyletic (aphyletic según Williams and Ebach, 2020) circumscriptions of the genera *Calandrinia* Kunth [s. l.] and *Talinum* Adans. [s. l.] had been diagnosed on the basis of the apparent sepals being persistent in the former and deciduous in the latter.[2] Hence, I attempted to include in my review of the taxonomy of the annual species of *Calandrinia* s. str. a brief discussion of the sepal fate trait. But this soon got out of hand, so I spun off this discussion into a separate work.

Briefly, reexamination of the perianth fate in the light of current Portulacineae taxonomic concepts revealed a previously unappreciated pattern. Unequivocally persistent sepals are characteristic of almost all Montiaceae, which are sister to remaining Portulacineae. An exception is the genus *Phemeranthus* Raf. This is notable because this genus possibly is sister to remaining Montiaceae, depending upon the relations of the poorly known and possibly extinct *Schreiteria* Carolin (see Hershkovitz, 2019a), which evidently has persistent sepals. Some *Phemeranthus* species are characterized as having persistent sepals, but evidently, unlike other Montiaceae, they are deciduous *eventually*. Deciduous sepals evidently evolved also in a few species of *Rumicastrum* Ulbrich, a relatively derived genus in the family.

Remaining Portulacineae include a well-supported clade comprising Anacampserotaceae, Cactaceae, Portulacaceae, and Talinaceae (the ACPT clade; Nyffeler and Eggli, 2010; all subsequent analyses). Sepals in all but Cactaceae are deciduous, but sometimes tardily so. Cactaceae evidently are nested among the other three families, although the exact relations of the latter remain unresolved. However, sepal fate per se cannot be evaluated for Cactaceae, because they have a multiseriate perianth consisting of more or less intergrading bracteate, sepaloid, and petaloid organs. However, among some Cactaceae, including some "basal" lineages, the *entire* perianth is deciduous.

The preceding suggests the possibility that the difference in sepal fate is indeed a diagnostic trait among Portulacineae. It distinguishes not *Calandrinia* s. l. from *Talinum* s. l., but rather most Montiaceae from the ACPT clade. However, the evolutionary origin of deciduous sepals in the latter cannot be established. Relationships among the intervening taxa, Basellaceae, Didiereaceae, and Halophytaceae, remain unresolved. Moreover, their sepal fate state also is ambiguous, as elaborated in the companion work.

Because the interpretation of perianth fate evolution among Portulacineae hinges on phylogenetic resolution, I intended to include in the companion paper a concise review of the latter. This also got out of hand, partially because of the discordance among molecular/genomic investigations and partially owing to my own peculiar ideas on phylogenetic reconstruction, these in turn consequent to my biological epistemology. These complexities necessitated spinning off the

[1] Most thus far still in preprint form, but these all have copyrights, ISSNs, and DOIs.

[2] Portulacineae sepals evidently are homologous with outgroup involucral bracts, and the apparent petals with outgroup sepals; Ronse de Craene, 2013). The immediate outgroup comprises Molluginaceae as currently circumscribed (Hernández-Ledesma et al., 2015; Ogburn and Edwards, 2015; all subsequent analyses).

present discussion of Portulacineae phylogeny[3] from the discussion of Portulacineae perianth fate evolution…itself a spinoff from a review of the taxonomy of the annual species of *Calandrinia*.

A. A "backbone" of Portulacineae phylogeny

Figure 1 presents an artificially linearized more-or-less consensus "backbone" phylogeny of Portulacineae *taxa* based on analyses of targeted gene and phylogenomic DNA sequence data.[4] The figure is taken from a companion work in preparation on the evolution of perianth fate. The backbone emphasizes nodes relevant to the evolution of perianth fate traits. Resolution at individual nodes is discussed in later sections. But I present this tree first as a platform for the following section, which addresses questions of the theoretical and empirical bases for phylogenetic inference. Some nodes are illustrated as unresolved, here intended to mean that proposed resolutions are not in agreement. It does not mean that the nodes represent true polytomies or are resolvable or irresolvable. Phylogenetic evidence pertinent to these nodes will be considered at the end of this work.

Figure 1 represents a synthesis based on but not limited to evidence from:

(1) whole genome DNA "baited" for 19 gene families considered to function in C4/CAM photosynthesis plus 52–53 other nuclear genes from 60 "Portullugo" taxa (Portulacineae plus Molluginaceae; Edwards and Ogburn, 2012) plus 11 outgroups (Moore et al., 2017, 2018);

(2) a similar analysis by Goolsby et al. (2018a, b), which added genome sequences from additional taxa (mainly *Rumicastrum* spp., misclassified there in *Calandrinia*; see Hershkovitz, 2020d) and also some transcriptome sequences, thus including a total of 156 Portullugo taxa plus 11 outgroups;

(3) ca. 8600 gene homolog clusters inferred from transcriptomes of 68 Portulacineae taxa plus 14 outgroups (Wang et al., 2018);

(4) cpDNA[5] whole genome sequences from 95 species of Caryophyllales, including 29 species of Portulacineae (Yao et al., 2019); and

(5) sequence data from two nuclear and five chloroplast loci[6] from among 5018 Caryophyllales taxon samples, including 1600 Portulacineae samples (Smith et al., 2018; cf. Smith et al., 2017).[7] The analysis provides a summary proxy for previously published targeted gene analyses (see below), but it

---

[3] Initially, the present work also included a summary of Portulacineae diversification age and historical geography. This, too, became cumbersome, thus is deferred to another spin-off.

[4] Figure 1 can be called, in the sense of Williams and Ebach (2020), a "generalized cladogram" inferred subjectively from "fundamental cladograms," or, given their methodological derivation, "phylogenetic phenograms." Williams and Ebach (2020) distinguished between the terms "cladogram" and "phylogenetic tree," though maintained that the latter necessarily are derived from the former. Since my perspective here is phylogenetic, I use the term "phylogeny" or "phylogenetic tree." I discuss later William and Ebach's (2020) perspective.

[5] Chloroplast DNA

[6] Nuclear ribosomal DNA (nrDNA) internal transcribed spacer (ITS) region; nuclear phytochrome C (*phyC*); chloroplast DNA (cpDNA) *matK*, *ndhF* and *rbcL* genes and *trnH-psbA* and *trnL-trnF* intergenic spacers. The combined average length of these loci is ca. 8000 base pairs (bp). The aligned length was 12713 bp.

[7] 1600 Portulacineae as follows: Anacampserotaceae (27), Basellaceae (7), Cactaceae (1247), Didiereaceae (19), Halophytaceae (1), Montiaceae (207), Portulacaceae (76), Talinaceae (16).

is problematic on account of incomplete sampling and sampling bias,[8] as well as alignment issues.[9] Despite the missing data and alignment inaccuracies, perhaps remarkably, the ML bootstrap analysis of the dataset (Smith et al., 2017) yielded a Portulacineae backbone consistent with resolutions derived elsewhere, though with reduced bootstrap proportions (BP). This observation will be discussed later.

Both the whole genome and whole transcriptome studies above undertook both ("rapid") ML analysis of concatenated matrices and multi-species coalescence (MSC) analysis of the individual loci. Smith et al. (2018) analyzed only a concatenated matrix. The concatenation approach is technically identical to that applied to maximum likelihood (ML) analysis of individual targeted gene analyses, in which nodal support is estimated using the nonparametric bootstrap. The MSC approach approximates the phylogeny based on "summary statistics" of the individual locus results. The approach bears similarity to "supertree" construction using "supermatrices" of trees derived from different datasets. A full (simultaneous) analytical multi-locus MSC phylogenetic reconstruction is considered to be computationally intractable (Cai et al., 2020). The transcriptome analyses (Wang et al., 2018) also conducted tests supposed to detect ancient horizontal gene transfer (HGT) events, but the results were not discussed in critical detail. The whole genome, whole transcriptome, and whole plastome studies analyzed complete and various partial data sets, the latter extracted according to various criteria of data matrix completeness and/or alignment quality.

In addition to the studies above, I consulted earlier genomic analyses cited in these works, as well as earlier analyses of various targeted gene sequences in various taxa, in particular Nyffeler (2007, cpDNA *matK* and *ndhF*, mitochondrial DNA *nad1*); Brockington et al. (2009; eight cpDNA loci including the entire inverted repeat plus nrDNA); Arakaki et al. (2011; ca. half of the cpDNA genome plus *phyC*); Anton et al. (2014; nine cpDNA loci plus nrDNA); Ogburn and Edwards (2015; four cpDNA loci plus nrDNA and *phyC*); and Ocampo and Mair-Sánchez (2018; five cpDNA loci).

I also examined a eudicot-level analysis of the single cpDNA gene *matK* (Lendel et al., 2013). This analysis focused on molecular dating of Cactaceae origins, hence also Portulacineae. The data are the same as for Arakaki et al. (2011; also published as a chapter in Lendel, 2013), but the topology differs somewhat. However, Lendel et al. (2013) did not discuss the topological differences between the two analyses.

---

[8] The authors did not articulate the following (which I extracted manually from Smith et al., 2017): The total 1600 Portulacineae samples included sequences of: ITS (633), *matK* (890), *ndhF* (145), *phyC* (170), *rbcL* (134), *trnH-psbA* (239), and *trnL-trnF* (495). However, sampling of each locus was taxonomically irregularly distributed among the families Anacampserotaceae, Basellaceae, Cactaceae, Didiereaceae, Halophytaceae, Montiaceae, Portulacaceae, and Talinaceae as follows: ITS (11, 3, 343, 7,1, 187, 75, 6), *matK* (19, 4, 764, 12, 1, 75, 7, 8), *ndhF* (6, 3, 9, 8, 1, 43, 68, 7), *phyC* (16, 2, 92, 17, 1, 24, 8, 10), *rbcL* (4, 3, 114, 4, 1, 4, 3, 1), *trnH-psbA* (0, 0, 237, 0, 0, 0, 1, 1), and *trnL-trnF* (0, 0, 489, 0, 0, 6 [only *Claytonia* and *Montia*], 0, 0). Sampling overlap is relatively low. For example, despite the greater sampling for ITS and *matK*, only 247/1600 taxa were sampled for both, and only 94/1600 (91 Cactaceae and 5 *Claytonia*/*Montia*) for both ITS and *trnL-trnF*. A grand total of TWO of 1600 taxa were sampled for all seven loci. The entire Caryophyllales data set, 5018 taxa, manifests similar or more extreme sampling incompleteness/bias.

[9] Smith et al. (2018) did not articulate their sequence alignment protocol. From their alignment (Smith et al., 2017), it appears to have been purely computational using arbitrary alignment parameters and without alignment edition, the gaps then treated as missing data in the phylogenetic analysis. The most variable loci have especially "stretched" alignments relative to actual sequence lengths. For example, *matK* is typically ca. 1650 bp, but the alignment stretches to 4243 bp (2.6X); *trnL-trnF*, typically ca. 960 bp, extends to 2117 bp (2.2X); and ITS, typically 580 bp, to 871 bp (1.5X). Across Caryophyllales, with increasing divergence, the "alignment" of the lengthy most variable portions of these loci appears to be essentially random. This cannot help but introduce error in the phylogenetic estimate.

The various DNA sequence analyses above agree with respect to most nodes of the Portulacineae backbone phylogeny. But their few disagreements loom large in character state reconstruction, including that of perianth fate traits. In order to explain these disagreements, the following section focuses on theory, especially on the ontology of DNA sequences and organisms and its implications for their evolutionary relationship. I return to interpretation of the Portulacineae data in the subsequent section.

B. Epistemological considerations in phylogenetic reconstruction

As noted, Figure 1 shows some of the nodes as unresolved. This is because different analyses have resolved them in different ways, and I do not favor any particular resolution. The term "conflict" has been applied to refer to *both* differently resolved organismal cladograms (e.g., using the same data but different methods) *and* to individual trait cladograms (including gene trees) that resolve organismal cladograms differently. I maintain that the latter instance is correctly termed "incongruence" rather than "conflict"[10] (Hershkovitz, 2019b). Lewis et al. (2016) formally defined both conflict and incongruence as "dissonance."[11]

In this section, I argue that apparent cladistic disagreements owe as much to theoretical misconceptions as they do to limitations of data or methods. The points reemerge in the subsequent discussion of evidence for Portulacineae relations. In this section, I emphasize the physically imperfect coupling between targeted gene and phylogenomic DNA sequence data with organismal evolution, as well as the imperfect coupling between prevalent theory and methods with evolutionary dynamics. I maintain that no established computational method, and possibly no conceivable method, *reliably* recovers or even converges upon true evolutionary history. Explicit and tacit mainstream assumptions to the contrary amount to superstitions or, alternatively, converge upon something *other* than true evolutionary history. Yet, paradoxically, evolutionary history is materially tangible, hence amenable to scientific enquiry.

I also suggest in this section that the disagreements in cladistic/phylogenetic interpretations may be a consequence of cognitive bias. In particular, the science of systematics might be especially sensitive to cognitive algorithms that tend to force resolution of that which is intrinsically irresolvable. But the forced resolutions themselves may vary, leading to disagreement. I follow this discussion with a recommendation that phylogenetics is best approached not from the perspective of resolution, but rather of reconciliation.

---

[10] Authors cited here (Moore et al., 2018; Wang et al. 2018; Yao et al., 2019) applied the term "conflict" to describe the discrepancies among gene trees relative to a taxon tree. "Conflict," "incongruence," and other terms have been considered synonyms, but incongruence refers to passive or incidental disagreement, discordance, or inconsistency, derived from Latin for "not meeting (or coming together)." "Conflict" is derived from Latin for "striking" as in head-to-head battle. In phylogenetics and science in general, "conflict" would describe competing hypotheses based on the same observations, each hypothesis ideally the "truth." The observations, in turn, may be incongruent, discordant, or inconsistent. This is the case applicable to traits in general, including multiple genetic loci, each of which has a history, but none of which can be considered in conflict with the organismal history.

[11] Later, Neupane et al. (& Lewis, 2019) defined dissonance as equivalent only to incongruence. I presume that phylogenetic dissonance would include, e.g., minor second and major seventh harmonic cladistic intervals. Alternatively, Lewis et al. (2016) possibly preferred "dissonance" over "incongruence" or "conflict," because the former could be represented mathematically by the symbol "*D*," whereas, in their mathematical formulation, the symbols "*I*" and "*C*" already were occupied by, respectively, "information" and "clade." An alternative solution would be to define "*D*" as "discordance," which is the same as incongruence. Or it may be defined as "dissidence."

1. Phylogeny occurs at multiple hierarchical levels

The current phylogenetic paradigm idealizes[12] evolution as a unified process in which organisms and "their characters" coevolve in unison. In addition, phylogeny usually is idealized as a process of successive *divergence*,[13] such that it can be represented as a fully bifurcating cladogram.[14] These idealizations together anticipate bifurcate character cladograms perfectly coincident with the species cladogram. Perhaps in deference to this expectation, mainstream praxis, first, (subliminally or deliberately) selects traits optimized for tree-like representation (viz. as a 2-dimensional, or 2-D data matrix), and then secondly, idealizes the *minimization* of discrepancy (by one or another standard) between the data and the optimized cladogram. The discrepancies then are swept under the rug. In practice, the expectation of full congruence expectation *never* is realized. Hershkovitz (2018a, b; 2019a, b; 2021b) sought to derive an evolutionary conceptualization that better explains both the recalcitrance of trait data to conform to 2-D matrices and the failure of so-biased data to conform to a single cladogram.

Hershkovitz (2019a, b; cf. 2021b) emphasized *epistemological* errors in the mainstream paradigm. For example, the mainstream paradigm has recognized for decades that "gene trees are not (organismal) species trees." But the meaning of this aphorism has not been understood properly. It has been acknowledged only in circumstances in which gene trees are manifestly incongruent with presumed organismal trees and/or which each other. The distinction generally is ignored in circumstances where trees appear to be congruent. But as Hershkovitz (2019a, b) emphasized, gene trees *never* can be organism trees, because they are ontologically distinct, even when they appear to be congruent. Moreover, different gene trees also are ontologically as distinct as are trees of different organisms.[15] The evolution of these distinct entities, however constrained to coevolve, is nonetheless decoupled. This demonstrates that evolution is not and cannot be a unified process, and that it should not be so idealized.

As emphasized in Hershkovitz (2019a, b), thermodynamic systems, including biological systems, are organized hierarchically. Processes at a particular hierarchical level proceed somewhat independently, except as constrained by processes at adjacent hierarchical levels, viz. upper and lower bounds. The historical decoupling of evolution at successive hierarchical material levels (from DNA sequence sites to loci to genomes to organisms) is predicted theoretically and proven empirically. It is thus a well-established biological fact.

Most molecular/genomic analyses, including those cited in this work, mention some (but never all) expected discrepancies between gene and organismal genealogy, e.g., owing to horizontal gene transfer (HGT, usually attributed to hybridization) and/or incomplete (gene/locus) lineage sorting (ILS; e.g., Moore et al., 2017, 2018; Wang et al., 2018; Yao et al., 2019). HGT pertains to a broader

[12] By idealize, I mean in the best of all worlds. Certainly the paradigm recognizes that this ideal is not real.
[13] In modern phylogenetics, this perhaps owes especially to Darwin's "Principle of Divergence" (see analysis in Richards, 2012). Hershkovitz (2019b) asserted that Darwin's cladistic conception of evolution was "part and parcel" of his theory of Natural Selection. This undocumented conclusion was deduced by casual reading of Darwin (1859). I had not read Richards (2012), but this provides the documentation.
[14] Or, alternatively, a fully bifurcating "cladistic parameter" sensu Williams and Ebach (2020).
[15] The mainstream phylogenetics paradigm has for decades attributed gene tree incongruence (relative to a given taxon tree) to loci having different histories (e.g., HGT or ILS). This underscores their ontological distinction, which, in turn, demonstrates that loci can *never* have the *same* history. The history of loci having congruent trees is, at best, *coincident* rather than the same. Clearly the mutations at one locus are not the same as those at another.

phenomenon, *replicative recombination*, which includes also *infra*genomic recombinatorial phenomena, such as classical chromosomal crossing over and also gene conversion. The ubiquity and potency of these processes (not even to mention transposable elements, or TEs[16]) is evidenced by the substantial homogenization of the thousands of copies of nrDNA genes (see below). These processes can be conceived of as "vertical gene transfer," or VGT, a complement of HGT. Just as ILS and HGT, VGT can yield phylogenetic incongruencies among analyzed loci. In particular, VGT recombines sequences of supposedly divergent and monophyletic gene family orthologs. This is significant, because VGT seems to have been substantially overlooked in phylogenomic analyses, which concentrate on VGT-prone gene family data.

Setting aside VGT, the current phylogenomics paradigm at least recognizes HGT and ILS as explanations for empirically inferred incongruencies. Yet, current praxis still *idealizes* phylogeny as a unitary and bifurcating and rather than a multilevel and sometimes reticulating decoupled process. It oversimplifies genome evolutionary dynamics to the point of distorting the very relations it predicates to resolve.[17]

At the root of the phylogenetics problem (no pun intended) is a notion attributed to Darwin and idealized by phylogeneticists, that of a single *exclusive* common organismal ancestor of members of a clade. The idealization is inherent in bifurcate phylogenetic trees (derived from bifurcate cladograms) in which nodes are represented as dimensionless points (Hershkovitz, 2019b). This ideal can be real only in the case of clonal organisms. But this is somewhat of a red herring. At the "population" level, hypothetical ancestors have many asymptotically genetically identical relatives. In the minimal (and never mind the actual) time necessary to have evolved modern contemporary distinct/individuated lineages, the possibility to distinguish the ancestor from its close relatives – perhaps ancestors of other modern lineages – approaches nil. Sexual reproduction and hybridization complicates matters further. Thus, even under the best of circumstances, the correspondence of phylogeny to bifurcate cladograms and the historical *reconstruction* of cladograms remain distinct matters.

Correspondence between phylogeny and bifurcate cladograms aside, it must not be forgotten that the "evidence" for phylogenetic relations is not *direct* observation of organismal genealogy. The evidence is *indirect*. It derives from observed/inferred similarities/differences among presumptively individuated and uncorrelated *traits* (cf. Williams and Ebach, 2020). These may be phenotypes or genotypes, including individual DNA sequence bases. Phylogeny is then deduced from the assumption that the shared traits of a cladogram are inherited genetically.

Note that the heritable trait sample is used to *proxy* for the organism.[18] But not even all heritable traits together can be ontologically equivalent to organisms. This should be clear from organismal death, the moment when the "organism" ceases to be. Yet, the organismal traits persist at least transitorily.[19] As emphasized in Hershkovitz (2019b), the individuated organism occupies the hierarchical dimension within a thermodynamic milieu that is tautologically defined as that pertaining

[16] For example, Vaschetto and Ortiz (2019) review evidence for TE-mediated evolution of nrDNA, the latter the most widely exploited molecular phylogenetic marker. I have not reviewed the recent literature a propos TE-mediated evolution across the whole genome. But such evidence improbably would emerge from current phylogenomic approaches that harvest only "protein-coding" genes.

[17] Williams and Ebach (2020) also asserted that mainstream cladistic/phylogenetic methods distort relations, but for very different reasons, as discussed later.

[18] "Fundamental" cladogram (sensu Williams and Ebach, 2020) terminals conventionally are labeled only with names of taxa rather than taxon-{trait set}. This possibly reinforces a notion that the analyzed traits and the organisms are the same ontologically.

[19] It is notable in this regard that many of the organs of "dead" humans can be transplanted to other humans, and they continue to function for decades.

to itself. This dimension is defined by its unique property of being "alive." Life itself is not a material entity, but a *process* that transcends the material properties of its traits. The quality of life of biological organisms is embodied in their autopoiesis (Maturana and Varela, 1992.) and determinism (enactivism; Varela et al., 1992; cf. Virgo, 2019).

Heritable traits, in contrast, are individuated entities populating *other* hierarchical dimensions linked to the organism by transitory physical material adjacency. Some heritable traits may be purely physical entities, e.g., a particular DNA base. Others may have life-like properties, such as replication and development. Some may themselves be organisms (e.g., heritable microbial associations) or organism-derived (e.g., chloroplasts). And heritable traits may exist at the supraorganismic level, e.g., their ecological associations and geographic distributions. Here, we enter the realm of the "extended phenotype," which might be thought of as traits that are phylogenetically conserved, yet not "coded in the DNA."[20]

The mainstream paradigm seems to idealize a fixed relationship between organisms and their traits. This is manifest in different contexts, from taxonomic keys/descriptions to character coding in both morphological and molecular cladistic/phylogenetic analyses (the latter often using a single individual samples per taxon) to current "phylogenetic comparative analyses" (PhCA) that code species "climate niches" according to the mean annual temperature and precipitation at the geographic locality of one to a few collections.[21] The idealization probably is reinforced by negligent oversight and/or ignoring variability revealed via scrutiny.[22] Hierarchy theory avoids this idealization by emphasizing that organismal traits at all hierarchical levels are free to vary within the tolerance limits imposed at their respective upper and lower hierarchical bounds. But within the tolerance limits, organisms and traits exist independently (i.e., have independent, albeit constrained, histories).

As trivial as the preceding might seem, it is precisely the *conditional* (as opposed to rigid) interdependence of trait and organismal evolution that *theoretically* prohibits reconstruction of the history of the latter on the *basis* of the former. Linking the two histories requires a lemma or

---

[20] The boundary between the organism phenotype and extended phenotype presents a philosophical dilemma, e.g., to which side shall we defer a spider's web, a bird's nest, *or even a 3-bedroom bungalow and the SUV parked in the driveway?* Both hierarchy theory and autopoiesis provide illuminating perspectives. Per hierarchy theory, DNA "encodes" (at best!) *nothing but itself*. It is merely a substrate with which other biological levels *interact*. Note that the relations between genome, transcriptome, and proteome (and beyond) are not one-to-one, and that *most* of the eukaryote genome is *not* expressed and therefore encodes "nothing." The hierarchical level of the whole organism behaves (or enacts) in a manner such that *supraorganismal* materials ("food," if you will) eventually channel towards the *reconstruction* of the DNA. This reflects the fundamental process of autopoiesis. Autopoiesis (a.k.a. "life") is defined not materially, but *operationally*, viz. by its consummation. It is not nearly mechanically rigid or faithful as reductionist dogma would suggest, which precisely explains DNA *evolution* (via Natural Drift; Maturana and Mpodozis, 2000). Moreover, owing to any of myriad factors that mitigate the life process, DNA reconstruction eventually inevitably *fails*. This precisely explains the Malthusian paradox. This renders as red herrings *both* the notion that DNA "encodes" organismal structure and behavior (any more than it encodes the bungalow or SUV) *and* the notion of the physical organism/environment boundary. Unfortunately, mainstream "modern" biological paradigms, including those of systematics, ecology, and evolution, base on obsolete reductionist Newtonian mechanics, thereby propagating explanations inherently false.

[21] Hershkovitz (2019a, 2020a) criticized the use of single/few locality data (e.g., Ogburn and Edwards, 2015; Smith et al., 2018) as proxies for widespread species; also, Hershkovitz (2020a) emphasized that *mean* climate data is misleading because it does not represent the actual temperature/moisture in the organismal milieu and does not capture the *tolerance limits* consequent to either or both of interannual climate variation and ecological context.

[22] Hershkovitz (2021b) emphasized rampant interspecific DNA sequence variation at supposed "species barcode" loci; the infraspecific divergences often are as larger or larger than interspecific divergences.

"background knowledge." But despite its mathematical/computational/technological sophistication, the modern phylogenetics/phylogenomics continues to *idealize* a lemma known to be false, viz. perfect coupling of trait and organismal history. To underscore this point, the paradigm has applied the term "conflict" in cases where trait and organism histories actually *agree* with theoretical expectations, as in the case of hybridization (cf. Hershkovitz, 2019b).

Underscoring the preceding, in the context of MSC, Rosenberg (2013) and earlier works described a region in phylogenetic tree space where multiple lineages diverge in a timeframe shorter than the average gene coalescent unit. In this region, (unrealistically) assuming random gene assortment, *most* gene phylogenies will incongrue[23] with the organismal phylogeny. Adding gene locus data will cause MSC analyses to converge on the wrong organismal phylogeny. Linkem et al. (2014) provided an empirical corroboration. More importantly, however, the confounding region in tree space has been termed the "*anomalous* zone" and the discordant gene trees as "*anomalous* gene trees."

Hershkovitz' (2019b, 2021b) criticism of the phylogenetics/phylogenomics paradigm overlooked discussion of the "anomalous zone." Here, I note that it represents both a generalizable *technical* proof of phylogenetic irresolvability under the (unrealistic) model idealized by the mainstream paradigm…*and* evidence that the paradigm indeed idealizes an erroneous model. This is because phylogenetic *diversification* has two prerequisites: heritable variability *and* sibling relationships. But these conditions guarantee ILS. More importantly, it underscores the errant idealization of phylogeny as a bifurcating tree. Only in this sense does the "anomalous zone" yield "anomalies." Biologically, per hierarchy theory, it is *not* an "anomaly."

Unfortunately, the phylogenetics paradigm, if not explicitly *presuming* phylogenetic resolvability, clearly *projects* that phylogeny is asymptotically resolvable, viz. via increased data/sampling. In fact, the journal *Molecular Phylogenetics and Evolution* (Elsevier, Inc.) predicates to be "dedicated to bringing Darwin's dream within grasp – to have fairly true genealogical trees of each great kingdom of Nature."[24] Darwin's "dream" was innovative, but technically it is incorrect (Hershkovitz, 2019b).

However, the notion of phylogenetic irresolvability does not seem to have been lost completely upon all phylogenomicists. Cai et al. (2020) concluded their work on Malpighiales phylogenomics by remarking that their methodological approach would be "*useful…to highlight instances where relationships are better modeled as a network rather than a bifurcating tree.*" Meanwhile, Hipp et al. (2020) concluded their work on oak phylogenomics with the rhetorical question "*how do we interpret the [cladistic] history…if it is really a collection of diverse histories…reflecting different evolutionary pathways, all equally real*?" And to severely edit and paraphrase Linkem et al. (2014), "*The lower…*[cladistic resolution]*…provided by…*[MSC]*…tree inference is potentially a more accurate reflection of the…*[phylogeny]*...Instead of marginalizing the…results, we should acknowledge that they are a likely consequence of…*[poorly resolvable]*…speciation history…*".

But even in these cases, the mainstream phylogenetics paradigm seems to be at best hitting green but not sinking the putt, or perhaps even seeing the hole. The suggestions of Hipp et al. (2020) and Cai et al. (2020) are consequent not so much to theory per se as they are to empirical failure of both massive DNA data and computational sophistication to resolve phylogeny. In other words, they emerge as "excuses." Even then, their conclusions manifest an air of tentativeness, perhaps necessarily so in deference to the reductionist tree-thinking mentality entrenched in mainstream academic and

[23] This word is obsolete, but nevertheless is the correct verb form associated with "incongruence."
[24] See https://www.journals.elsevier.com/molecular-phylogenetics-and-evolution

editorial (hence also funding) institutions (see Hershkovitz, 2018b). In contrast, the conclusions of Hershkovitz (2019a, b, 2021b) are not minced. They are derived logically on the basis of solid biological and mathematical theory.

2. Evolution is a nonstochastic process

Current phylogenetic/phylogenomic praxis models evolution as a *linear stochastic* process. It is critical to recognize that the purpose of the model is not precisely to reconstruct evolution, but to *predict* it, albeit retroactively. In other words, the reconstruction is the history that best matches empirical observation with predictions of the model. If the model is correct, ML analysis of *increasing data* should *converge* on the correct history.

But as emphasized by Hershkovitz (2019b, 2021b), the generalized ML phylogenetic model is biologically implausible. Evolution at all hierarchical levels is *idiosyncratic*, driven autonomously by organisms and reflecting both stochastic and formally *chaos-like* processes. Chaotic processes can be predicted/reconstructed exactly only if the function and starting conditions are known. Lacking these, they cannot be predicted or reconstructed statistically. But whatever controversy this statement might provoke, there can be no controversy that evolution axiomatically overwrites evidence for itself, such that phylogeny itself becomes asymptomatically irresolvable, even in the case of stochastic evolution.

To review from Hershkovitz (2019b, 2021b), the chaos-like component of evolutionary idiosyncraticity is inherent in the *determinism* of organisms, in turn inherent in their defining quality, autopoiesis. This determinism distinguishes biological organisms from strictly physicochemical nonliving entities. Thus, the behavior of organisms is conceived properly as *enaction* rather than *reaction* (Varela et al., 1992; cf. Virgo, 2019). Within their limits of tolerance to environmental conditions at adjacent (higher/external and lower/internal) hierarchical levels, organisms have freedom to determine, or enact, their history. This freedom renders unpredictable their enaction. Determinism and unpredictability render a chaotic quality to biological behavior and evolution. However, the chaotic quality indeed is mitigated by conditions at adjacent hierarchical levels, and these conditions collectively behave indeterminately/stochastically. Thus, Hershkovitz (2019b, 2021b) formally described evolutionary idiosyncraticity (PEI)[25] as evolutionary behavior consequent to *both* determinate/chaotic and indeterminate/stochastic functions.

A corollary of PEI is that the evolutionary process reflects an "*infinite order*" Markov process. In other words, evolutionary transitions are conditioned by *all* historical states since process initiation. This is consequent to the chaos-like component of idiosyncraticity. Meanwhile, the current mainstream paradigm first *idealizes* and then *correspondingly* models evolution as a *first-order* Markov process, in which the evolutionary transitions are conditioned by current and not prior states.[26] It must be stressed that the erroneous assumption of a first-order process precisely justifies, or at least renders tractable, inductive (statistical) evolutionary analysis.

---

[25] Hershkovitz (2019b, 2021b) dubbed this as the "Principle of Evolutionary Idiosyncraticity" (PEI), summarized as **f**(chaos)**f**(stochasticity), or the "God function."

[26] Current praxis commonly conditions evolutionary (retroactive) transition probabilities on empirically estimated "global" history, e.g., in base substitution models. However, "global" history is not the same as "prior" history. The "global" history is estimated from the history of *each* analyzed taxon. Thus, as the number of analyzed taxa and the divergence between them increases, the less relevant is the "global" history to the true history of any one of them. "Global" history is a lie.

Intuitively, the preceding should not be difficult to grasp. A notable feature of the massive accumulated biodiversity literature is the focus of individual analyses on particular taxa, ranging from single species to groups of related species up to and including major phyla, e.g., plants or animals. It is accepted at least tacitly and often explicitly that taxa have *distinct/divergent* evolutionary/ecological trajectories. Targeted-gene molecular phylogenetic analyses are undertaken using loci that are shared and both sufficiently conserved (viz. alignable) and variable among members of the circumscribed *clade of interest*. Which is to say that, in the universe of genetic loci, remarkably few lend themselves to analysis of relations among *any particular taxa*. The same is true of phylogenetic analyses and PhCA of phenotypes, e.g., reproductive morphology of particular plant or animal clades.[27]

But here is another case where the forest is missed for the (phylogenetic) trees (cf. Hershkovitz, 2019b). Current praxis seems to blind investigators from an understanding that taxa "of interest," descended from a common ancestor that is *excluded* from the analysis and therefore *not* "of interest." The fact that genotypes and phenotypes have taxonomically restricted ranges of analytical utility or applicability itself is evidence that traits of each successive common ancestor irreversibly influence the traits of all of its descendents. This is the same as saying that evolution approximates an "infinite-order Markov process," which is the same as a chaotic process.

PEI is not only predicted theoretically, it is manifest in essentially all empirical data analyses. But it has been overlooked. For example, reconsideration of the "ancient" paradigm of phenotype-based phylogenetics[28] reveals not only how few are the data that lend themselves to linear interpretation, but how many of these few in reality do not. Despite the phenotypic complexity of organisms, generally systematists were challenged to identify a number of discrete, phylogenetically informative characters/character states sufficient to resolve phylogenies with any degree of mathematical credibility (e.g., bootstrap support) or even at all (i.e., single most parsimonious, or MP resolutions).

But phenotype-based phylogenetics were hampered as much by the quality of characters as the quantity. The systematics literature of the period is replete with collectively inconclusive discussions of how to define and code phenotypic data objectively, without mathematical bias, and without introducing illogical evolutionary transitions.[29] Implementing these suggestions in practice was no less

[27] As Goloboff et al. (2018: 625) remarked, "The homogeneous Markov model assumes (when applied to morphology) that the characters are like units that can simply switch into one or another state at any point in the tree…But the very fact that taxonomists beginning to investigate a group first need to learn the relevant characters speaks against the idea that all groups can be classified by looking at the same sets of characters randomly changing over all the tree; that is why someone who has worked extensively on spider morphology needs to learn a whole new suite of anatomical characters if starting now to work on, say, beetles."

[28] The phenotype-based phylogenetics bandwagon developed in the 1970s mainly in zoology, and then infiltrated botany in the 1980s. It peaked in the late 1980s to earliest 1990s. Its upstream "promoter" was not so much theoretical developments (dating to the 1960s), but the development of cheap and user-friendly personal computers and phylogenetics software. The abrupt "stop codon" was the development of PCR and reasonably rapid DNA sequencing.

[29] Conundrums included but were not limited to: the distinction between characters and character states, especially in the case of presumed serial homologues; a priori ordering of presumed multistate characters (e.g., numbers of chromosomes or floral parts); definition/scoring of polymorphic or continuously variable characters or those representing states of developmental pathways, including embryonic characters and secondary metabolites; functionally and/or developmentally "correlated" characters; characters "missing" in some taxa (hence illogically have "no" character state); a priori and a posteriori character weighting; and criteria for cladogram optimization.

challenging. In the end, subjective bias unavoidably influenced analytical results.[30] This was not necessarily a problem if the bias favored the "true" phylogeny.[31] But this exposes as somewhat farcical the deference to reductionist mathematical objectivity.

In retrospect, it can be appreciated that the difficulties encountered in phenotype-based phylogenetics owes to the reductionist misconceptualization of organisms, hence their evolution. It does not owe per se to the conceptualization of clades, which are tangible, hence discoverable, entities. But the reductionist approach canalizes an expectation that phenotypic characters are merely independently evolving *components* of organisms. This translates into an expectation that they *should* assort neatly into a 2-D data matrix.[32] Accordingly, theoretical and empirical efforts attempted to force-fit data into this framework.

Appreciation for PEI and its pillars (especially autopoiesis) would have preempted reductionist expectations. For example, PEI anticipates ambiguity/ambivalence in character definition and coding, i.e., which taxa are scored as being the same. Coding similarity, in turn, canalizes analytical results. PEI predicts the observation that characters coded as being the same nonetheless manifest subtle to striking taxon-specific differences. This, along with autapomorphies, is consequent to evolutionary *expansion/contraction* of character space evolution, rather than merely oscillation within a fixed space. But perhaps most importantly, PEI predicates a chaotic component in the evolutionary process. This chaotic component *predicts* character space expansion/contraction. This, in turn, falsifies any notion of (strictly) mathematical/statistical reconstruction, unless the evolutionary function and initial conditions are known.

DNA-based phylogenetics was supposed to have overcome both the paucity of informative phenotypic traits *and* the subjectivity of phenotypic interpretation. But the supposed objectivity is illusory. If phenotypes mislead because of their intergrading complexity, DNA bases mislead because of their stark simplicity. Nothing about the *structure* of a DNA base can distinguish between homology (i.e., synapomorphy) and convergent evolution at a presumed homologous (sensu Williams and Ebach, 2020; viz. aligned) DNA sequence site.

---

[30] A propos Portulacineae phylogeny, some aspects of phenotypic character analysis are discussed in Hershkovitz (1989, 1993). The former reconsidered the incompetence-mediated analysis of Rodman et al. (1984), and latter reconsidered the more illuminating character analysis of Carolin (1987).

[31] An example is Hershkovitz' (1993) coding of the inflorescence as solitary/axillary in both *Calandrinia* sect. *Caespitosa* Phil. (as *C.* sect. *Acaules* Reiche) and *Calandrinia* sect. *Calandrinia*, whereas Carolin (1987) scored the former as solitary/axillary and the latter as cincinnoid. *Calandrinia* Kunth (sensu Hershkovitz, 1993, 2019a) was polyphyletic in Carolin (1987; cf. Carolin, 1993), but monophyletic in Hershkovitz (1993). The scoring in Hershkovitz (1993) was influenced partially on a different subjective morphological interpretation and partially on a biased opinion based on *other* morphological evidence that *Calandrinia* sensu Hershkovitz was monophyletic. Its monophyly later was corroborated by DNA evidence.

[32] DNA sequences abruptly marginalized phenotypic characters in phylogenetic analysis. But comparable phenotypic character analysis resurged in the paradigm of macroevolutionary ecology using PhCA. While phylogenetic analysis seeks to identify phenotypes that are homologies (viz. diagnose monophyletic lineages), PhCA seeks to explain, in a Darwinian framework, *homoplasious* and therefore merely similar phenotypes shared among otherwise historically independent (aphyletic sensu Williams and Ebach, 2020) lineages. Obviously this violates explicitly the very premise of statistical analysis, in which sampled individuals are idealized as ontologically and historically *identical* except for differential "treatments." Hershkovitz (2018b, 2019a, b) criticized and rejected the Darwinian PhCA paradigm in general and in relation to empirical analyses of Portulacineae phenotypes. Classical questions of character ontology (viz. definition and coding) emerged in the latter. However, the criticism focused not on this, but on the fallacies of reductionist Darwinian dogma, both theoretical and as betrayed by empirical application. Darwinian explanations for character correlations fail the criteria of necessity and sufficiency. And because of idiosyncraticity, they are patently false.

In fact, despite prevailing belief to the contrary, a shared DNA base, by itself, *cannot be* a homology, no more than an organism's electrons and protons. In contrast to morphological structures, which evolved *during* phylogeny, all DNA bases *predate* the origin of organisms. They represent a homology at the boundary between the RNA and RNA/DNA worlds. The homologs in molecular systematics are contextual, referring to other hierarchical levels, e.g., a particular aligned sequence *position*. Meanwhile, just as in morphology, this context itself evolves. In this sense, the entire sequence represents a single character *suite* comparable to, e.g., the vertebrate skeleton or the angiosperm flower.[33] When the sequence is conceived in this way, the complexity of molecular evolution becomes more apparent, and the notion of stochastic base substitution models ever more absurd.

Both phenotype-based and DNA-based analysis can be highly contrived. The "phylogenetic utility" of DNA sequences derives substantially from sanitizing the genetic data so that they can be force-fit into the conventional analytical paradigm. In particular, genetic loci are selected taxon-specifically for alignability and modest variability[34] qualities, and then the cherry-picked DNA sequences are trimmed to minimize/eliminate their length variability. This creates the illusion that DNA sequence evolution is nothing more than relatively few "independent and identically distributed" base substitutions that can be counted easily and analyzed statistically. The illusion is not merely reinforced by the contemporary paradigm, it was deliberately created by the latter, because substitution *quantity* is really all the contemporary paradigm can deal with. Standardization of the praxis and protocols reinforces and illusion of objectivity. More consequentially, in order to maximize "phylogenetic information," the paradigm deliberately distorts the *molecular evolutionary* information.

Indeed, even within artificially homogenized data, it has been determined both logically and empirically that the symmetric (i.e., fully reversible) substitution weighting schemes and statistical models[35] incorporated in prevailing analytical methods fail to capture the complexity of their

[33] Although my thoughts here on this matter were conceived independently, later I located similar sentiments, which is gratifying. Williams and Ebach (2020: 199) had mused, "…what are the homologues? Are they the base pairs, the genes, or both?" They go on to cite relevant opinions by Springer and Gatesy and also quote Inkpen and Doolittle (2016): "In molecular phylogenetics…..we must contextualize homology if we are to talk about it at all." My apparent reinvention of this wheel owes not to these authors, but only to my scrutiny of the Portulacineae molecular/genomic studies. These render an impression that, in empirical research practice, indeed it is *not* talked about at all.

[34] Variability across a "good" data set is on the order of no more than 10% base differences. Data sets with higher extremes tend to enter the realm of both mutational "saturation" at individual sites and, independently, problematic length variability.

[35] The current and most widely applied statistical model scheme discriminates up to 12 parameters: up to six substitution types, four base composition categories, and two types of among-site rate variation. (Usually the program "default" of four gamma rate categories is used, each of which might be considered an additional parameter. Theoretically, the number of rate categories could be as high as the number of sites in the alignment. But I cannot recall ever coming across an empirical analysis that applied and justified a number of rate categories different from the default.) Bayesian methodology may permit additional discrimination between sequence data partitions (viz. base positions or loci presumed to share the same phylogenetic history). Needless to say, the greater is the parameterization, the lesser the statistical power. But in practice, statistical power often maxes out with *fewer* than the standard 12 parameters. And even though the standard 12 parameters often already exceed the number statistically resolvable in typical data sets, theoretical and empirical evidence suggests that this number is a severe *underestimate*. Current models assume that *individual* sites evolve independently, ignoring evidence of molecular coevolution, such that evolution at particular sites may be conditioned by factors operating directly or indirectly both in cis (other sites in the same locus, including adjacent sites) and in trans (sites in other loci), as well as upstream from replication and downstream from

molecular evolution (Hershkovitz, 2021b).[36] DNA sequences do not evolve uniformly –substitution quality and/or quantity *vary* across the phylogenetic tree. Notably, they vary at different *hierarchical* levels, viz., at all sites, among sites, among loci, and/or among lineages. But the *significance* of these observations seems to be lost upon the most vaunted of reductionist molecular evolutionary theoreticians who *developed* the prevailing analytical methods, and whose very intelligence evidently is vastly overestimated. The methods themselves literally obscure visibility of idiosyncrasy in consequent empirical analysis.

Perhaps the most systematically suppressed clue to the inadequacy of the prevalent paradigm is evidence for nonstationary substitution dynamics (Hershkovitz, 2018b; cf. 2019b, 2021b).[37] The most conspicuous and ubiquitous evidence is marked difference in base composition among analyzed sequences (e.g., Hershkovitz, 1997, 2000, 2021b). This demonstrates that substitution dynamics cannot be constant across the data matrix.[38] Prevalent praxis force-fits "statistically significant" base compositional bias across a data set[39] into single average "equilibrium frequencies,"[40] which evidently and quite literally do not exist.[41]

The important observation here, however, is not whether or not the base frequencies or substitution model is correct. The critical point is that a fixed substitution model effectively assumes and enforces a fixed substitution "character space." Within this space, substitution is assumed to be stochastic. Nonstationarity demonstrates, firstly, that the substitution model estimates across all sequences in an alignment can never be accurate along any particular tree branch. More importantly, however, it demonstrates that substitution character space evolves continuously and unpredictably, in agreement with the PEI.

A more subtle clue for idiosyncraticity emerges in DNA sequence length variation. Length variation is consequent to processes[42] distinct from that of substitution, i.e., base misincorporation. At

---

transcription. Nothing articulated here is "new" or unusual. All of it can be gleaned from 1990s vintage literature, including *undergraduate* level molecular biology texts.

[36] Hershkovitz (2021b) aptly dubbed currently applied substitution models as "fake models." This reflects not their per se technical inaccuracy/inadequacy, but rather the ubiquitous failure of empirical analyses to acknowledge the rarely advertised theoretical *knowledge* of their inaccuracy/inadequacy. The term "fake" applies to predicating as "true" something reasonably evidently known to be "false."

[37] Current molecular dating procedures explicitly presume nonstationarity of overall evolutionary rate but otherwise are constrained to presume that the substitution model itself (estimated by averaging across the entire data set) has been constant.

[38] It is not appreciated that the stationarity effectively presumed to be "true" in empirical analyses is "methodologically imposed," viz. by the pooling of all base variability in an alignment into one or a few categories in order to contrive statistical power. Nonstationarity implicates multi-parameter branch-specific substitution dynamics, which are statistically impractical and likely computationally unfeasible. Thus, current praxis deliberately and systematically conceals nonstationarity. This, in turn, reinforces an illusion of stationarity.

[39] Hershkovitz and Zimmer (1996) and Hershkovitz (2021b) demonstrated marked differences in base composition among constant/conserved and variable sites. Current methods permit manual specification of "equilibrium" frequencies to accommodate such observations. In practice, however, empirical analysts apply automated model generation protocols that do not detect such differences.

[40] Conceptually, prevalent praxis presumes that base substitution probability is determined partially by the pool of available bases as manifested in the average.

[41] Likewise, base composition may be functionally constrained at particular sites, e.g., polyadenylated regions associated with transcription, which are common in cpDNA noncoding sequences exploited in phylogenetic analysis.

[42] E.g., replication misalignment involving sequence repeats or secondary structure; mobile element insertion/deletion.

the level of the nuclear genome, length variation, if not more prevalent than base substitution variation, is certainly more extreme. Even sometimes to the interspecific level, much of the nuclear genome cannot be meaningfully aligned or even parsed into homologous loci. Across all of life, the proportion is negligible.

As noted, targeted gene phylogenetic analysis focuses on the relatively few sequences that are highly alignable among the sample taxa,[43,44] and ignores the hypervariable and ambiguously or not alignable regions characteristic of the bulk of the genome. It seems perfectly logical in phylogenetic analysis to ignore characters that cannot be meaningfully compared. But these characters are not phylogenetically meaningless. Here expressed maximally is the evolutionary fluctuation in character space anticipated by PEI . Again, the contemporary paradigm systematically suppresses this evidence, biasing in favor of character evidence *apparently* more consistent with the stochasticity model.

Perceptualization of the entire genome[45] facilitates appreciation of its idiosyncratic evolution. Both base and length mutations are influenced by processes at various hierarchical levels, as specified formally by the PEI (cf. Hershkovitz, 2019a, b, 2021b). The genome perspective reveals the nonindependence among sequence characters, viz. how they influence each other in the genetic hierarchy.[46] In fact, the primary DNA structure has no particular grammar (much less text[47]). The primary structure indeed influences certain replication infidelities,[48] and these may establish evolutionary patterns, including the evolution of "loci." But most aspects of sequence evolution, including the parsing into the transitionally stable "genes" exploited in comparative analysis, are influenced by processes above and beyond the level of DNA primary structure,[49] up to and including ecological levels.

Thus, the evolution of the whole genome renders clear the farce of conventional stochastic substitution models (i.e., "fake models") hawked by the contemporary phylogenetics paradigm. The illusion of a uniform substitution process operating on a pool of independent, identical, and identically distributed DNA sequence bases is sleight of hand. As noted above, it is created by cherry-picking modestly variable and length-uniform loci and ignoring overwhelming contrary evidence from the overwhelming proportion of the genome.

---

[43] Most data sets do include indels. Sometimes these are incorporated into the analysis as separate discrete characters, but often the questions of homology, multistate character ordering, and/or reversibility emerge, the same as with phenotypic characters. Length variability thus is "ignored" in base substitution analysis, but it introduces the problem of "missing data," whose theoretical implications are beyond the scope of the present work and whose consequences are not understood by most empirical analysts. The methodological default option is to "ignore" missing data. Out of sight, out of mind. Williams and Ebach (2020) noted (correctly) that mainstream phylogenetic methods *cannot* "ignore" missing data. Effectively, they parse data that do not exist.

[44] As described above, Smith et al. (2018) arbitrarily aligned unalignable regions of some loci. The loci were substantially to perfectly alignable among the closely related taxa, but not across all of Caryophyllales.

[45] All regions, from the more conserved sequences exploited in phylogenetic analysis to the more prevalent variable regions that the phylogenetic paradigm ignores.

[46] For example, base substitutions can interrupt sequence repeats, modify DNA secondary structure, or alter mobile element insertion sequences, thus influencing length variation mutations, while length variation mutations influence subsequent base substitution evolution in myriad ways.

[47] Meaning structural demarcation into, e.g., genes and exons.

[48] These include the natural bias for substitution transitions over transversions and length variations owing to sequence repeats and DNA secondary structure.

[49] These influences on evolution are misconceived as Darwinian "natural selection," a model incompatible with idiosyncraticity (Hershkovitz, 2019b).

The preceding is not to say that "empirically estimated" models are not (technically) statistically superior to a random model. But this is a red herring, because it is well-known in statistics that a statistically superior model is not the same as a "true" model. Highly complex phenomena defy reduction into few parameters, such that an apparently statistically superior model may mislead empirical analysis and, contrapositively, an inferior model may perform better (Abadi et al., 2019). More theoretically problematic in this case, as emphasized by Hershkovitz (2021b), evolution is *never* random. It is *always* idiosyncratic.

3. Phylogenomics: more is better?

Because the present reevaluation of Portulacineae phylogeny emphasizes phylogenomic evidence, I review here the relationship between phylogenetic history and phylogenomic praxis,[50] viz. data and methods. Phylogenomic approaches have been hawked, in part but still very specifically, under the premise that "more is better," viz. that they facilitate resolution of phylogenetic nodes not resolved using fewer data. This is the "shock and awe" approach, bombarding the nodes with megatons of sequence data. But contrary to technocratic expectations – yet in agreement with now classical molecular evolutionary theory – phylogenomic analyses at best *refine* phylogenetic interpretation, but they do not seem to *resolve* cladograms per se much or at all better than earlier targeted gene analyses (cf. Hershkovitz, 2019b).

In fact, if both DNA sequences and organisms evolved as "ideally" as early molecular systematists had fantasized, then very little DNA sequence would be needed to resolve the "Tree of Life" (e.g., Soltis and Soltis, 2018). Notably, much of the *credible* resolution of said tree indeed was resolved with remarkably little sequence data, usually less, sometimes much less, than 10,000 bp. Sometimes such few data resolve analyzed clades relatively well.

For example, in a phylogenetic analysis of the sister genera *Chaetanthera* Ruiz & Pav. and *Oriastrum* Poepp. & Endl. (Asteraceae; Mutiseae), two loci[51]comprising just 1400 bp resolved 70% of the tree nodes with ≥ 70% BP (Hershkovitz, 2021b). But this was luck of the draw. As counterexample, South American Montiaceae were far less consensual to molecular systematic analysis (Hershkovitz, 2006). Using ITS and a comparable cpDNA locus, only about 20% of the 182 possible nodes were resolved with ≥ 70% BP. In any case, among all angiosperm analyses, probably no more than half of the nodes resolve using a combination of a nuclear and cpDNA locus.[52] And this precisely is why whole genome data are pursued.[53] Otherwise, the phylogenetic task would have been completed a decade or more ago.

---

[50] I refer here principally to the *nuclear* genome in plants and eukaryotes generally. In plants, the nuclear genome typically includes 2–6 million DNA base pairs. The chloroplast genome, 100–150 thousand base pairs, behaves more or less as a single nonrecombining locus, though the inverted repeat and adjacent sequence recombines. Structure and evolution of the plant mitochondrial genome (mtDNA) is too complex for most phylogenetic reconstruction purposes. This has been known since the 1990s and has been corroborated by the much later mtDNA analysis effort by Nyffeler (2007).

[51] ITS and the cpDNA *rpl32-trnL* spacer..

[52] Though in many cases the lack of resolution in combined data analyses reflects incongruence between the nuclear and cpDNA marker.

[53] Smith et al. (2018: 846) expressed the decades-old prevalent "more sampling, more data" credo: "Confident resolution of many of the systematic relationships will require genomic and transcriptomic sampling, as well as more thorough taxon sampling…". This is axiomatic, but their poor sequence alignment (Smith et al., 2017; see discussion above) suggests that the authors would have done better to analyze competently the data at hand.

But for many reasons, regardless of the cladistic resolution obtained by relatively short sequences, it is not clear that 100–1000 times as many bp can resolve cladograms marginally much better. Indeed, phylogenomic analyses demonstrate quite the contrary. As I have noted (Hershkovitz, 2019b), the targeted genes had been selected precisely because they optimized the criteria to be considered "reliable" phylogenetic markers.[54] This selection was somewhat analogous to primitive agriculture, selecting plants that were easiest to cultivate, had the best yields, and did not kill you. Laborious analyses and experience showed that, albeit always with exceptions, the targeted genes manifested the "most" predictable evolutionary behavior across a broad phylogenetic range, an evolution not inherently correlated with morphological phenotype or environment (i.e., "neutral" at least in these respects), and ease of discrimination from other loci (but see also below). In contrast, and metaphorically, phylogenomic pipeline praxis drives an International Harvester through uncultivated vegetation.

As I emphasized in Hershkovitz (2019b), the phylogenomic paradigm seems to have been conceived with a naïve idealization[55] of the genome as simply a very long DNA sequence whose evolution otherwise is not fundamentally different from that of conventional targeted genes. Thus, the praxis is to pool the data into a single analysis. Cladogram resolvability thus is conceived as a function of sample size, whether base positions or loci or both.

But since *most* of the genome would be considered suboptimal to useless for labor-intensive targeted gene analysis, it hardly would seem to be suitable for comparatively indiscriminant bulk analysis. As discussed below, this unsuitability owes in part to chromosomal-level evolutionary artifacts and partly to violation of the assumption of phylogenetically unbiased DNA sequence evolution. Axiomatically, increased locus sampling increases the likelihood of other analytically unmodeled evolutionary complexities. Thus, increasing data size disproportionately increases the possibility of error. The errors are not readily detected/corrected by automated analytical methods. This is partially because current praxis substitutes artificial for genuine intelligence, and partially because prevalent analytical methods derive from naïve conceptualizations of organisms and their evolution.

4. Phylogenomic gene tree incongruence – more than HGT and ILS?

Whether expected or not, one of the more striking outcomes of phylogenomic analyses is the degree of incongruence (historically often referred to as "conflict" and more recently as "dissonance;" see above) among individual locus gene trees with respect to any possible organismal tree. Such is the case with Portulacineae phylogenomic data (Moore et al., 2017, 2018; Wang et al., 2018). Typically, the incongruencies are presumed to be consequent to HGT and/or ILS.[56] While both HGT and ILS are

[54] This is not at all to say that such markers yield true species trees, since gene trees *never* are species trees; or that they even yield true *gene* trees. For example, nrDNA evolution is far more complex than assumed by molecular phylogenetic reconstruction methods (see., e.g., Symanová, 2019 and discussion below).

[55] I doubt that phylogenomicists are completely naïve in this respect. Phylogenomics literature, like that of molecular systematics before it and perhaps the reductionist scientific literature in general, manifests an unmistakable "marketing" quality, and naïve simplicity is what sells. This hardly is surprising given modern "competitive" scientific culture, in which the competition is much more socioeconomic than it is intellectual.

[56] Cai et al. (2020) analyzed gene trees incongruencies among 423 genomic loci from divergent members of the angiosperm order Malpighiales. The crown group lineages supposedly diverged ca. 100 million years ago and have been considered "recalcitrant" to cladistic resolution. Cai et al. (2020) estimated that one third of the incongruencies owed to "gene tree estimation error" (equivalent to irresolvability or "noise"), whereas at best 10% could be attributed to ILS and 21% to ancient HGT (hybridization). However, their model makes several inexplicit but unrealistic assumptions, among these, that the loci are orthologous and evolutionary independent and identically distributed. They did not consider VGT.

predicted to result in gene tree incongruencies, discriminating between these – not to mention other less advertised causes of incongruencies (VGT, mistaken orthology, and convergence; see below) – presents a puzzle that remains to be solved.

Apparently unnoticed in this discussion is how the "universal" HGT/ILS explanation becomes decreasingly plausible with increasing divergence. Subsequent evolution overwrites evidence for these events. At what divergence degree such evidence persists – and can be distinguished from homoplasy – is not clear. It would seem to be not very long. This is because divergence at ILS and HGT loci inherently is low, because the source organisms must be relatively closely related.[57] With evolutionary time, the vertical divergence at these loci eventually should exceed the horizontal divergence.

It has been suggested that mistaken inferences of gene orthology might yield gene tree incongruencies in phylogenomic studies (e.g., Walker et al., 2018). Phylogenomics approaches are less rigorous than labor-intensive classical molecular genetic approaches that probe the genome exhaustively for gene family members and then corroborate orthology inference with expression analysis. Phylogenomics praxis usually isolates genes only in bulk and then infers orthology purely computationally. The former procedure might fail to recover all gene family members and the latter, independent of the first, might sort paralogs incorrectly.

In the introduction to their Portulacineae phylogenomics analysis, Moore et al. (2018) suggested another source of mistaken orthology inference. They emphasized that so-called single-copy loci (SCLs; including those exploited in targeted gene analysis) may have had an evolutionary history of duplication (creating paralogs) followed by differential loss of one or the other copy in subsequent descendent lineages (i.e., paralog "birth-death"). This is logical considering that, historically, duplications and losses occur commonly from the levels of single bases to loci to whole genomes (Vaschetto and Ortiz, 2019). In this case, such birth-death could result in SCLs being paralogs rather than orthologs. This, along with hybridization, would mislead phylogenetic reconstruction. Here (with apologies to Williams and Ebach), I use the term "pseudologs" for such loci that appear for all the world to be perfect homologs when they are not.

Moore et al.'s (2018) point is well-taken. Both targeted gene and some phylogenomics analyses have emphasized the search for SCLs on the false assumption that these inherently are orthologs. Orthology is *one* of the desiderata (see below) for *optimizing* (though never ontologically attaining) congruence between gene trees and species trees. But Moore et al. (2018) did not explain why pseudology should be any *less* of a problem for *their* approach to phylogenetic reconstruction using highly duplicated *multi*-copy loci. I argue that the problem could be much worse and result not only in greater duplication/extinction of paralogs, but also greater VGT, hence mistaken orthology.

To appreciate the proceeding, consider the case of nrDNA. Ignoring TEs, these are normally the most duplicated functional genes in the nuclear genome. The genes usually are arranged, often at more than one chromosomal locus, as repeated copies of a large cistron that includes the 18–28S nrDNA

[57] This is not *always* the case (as idiosyncraticity would have it). The crown group (CG) of oaks (*Quercus*; Fagaceae) is on the order of 56 million years (MY) old and the sectional CGs ca. 25–48 MY old (Hipp et al., 2020). Species within sections notoriously hybridize (Kremer and Hipp, 2019), and fertile intersectional hybrids are known (Vázquez et al., 2015; Hipp et al., 2020). But oaks are perhaps the most extreme example given their unique combination of characteristics. The 400 species share the same diploid chromosome number (Denk et al., 2019). Also, oaks commonly live several hundred years. Thus, alleles can transmit not only horizontally, but "vertically" through time, skipping over 500 subsequent generations. And oaks are wind-pollinated, which maximizes allele dispersal within and among the essentially "incompletely formed" species. Thus, expected high levels of HGT and ILS predict, in turn, considerable gene tree incongruency. This is observed (Hipp et al., 2020), though additional processes (see below) must not be ruled out.

gene (or its homolog),[58] and the highly duplicated but usually unlinked 5S nrDNA gene.[59] NrDNA represents the classic case of "concerted evolution" of repetitive sequences, generally attributed to either or both of asymmetric chiasma or gene conversion consequent to interlocus pairing at replication (Escobar et al., 2011; Symonová, 2019; see also Harpke and Peterson, 2008), as well as locus birth-death (Zhang et al., 2021). The consequence is substantial homogenization of the 100–1000s of copies both within and among loci. This homogenization evidently can occur rapidly, even in the case of hybrids (e.g., Lunerová et al., 2017; see also Garcia et al., 2017). Above, I referred to this process as a form of VGT.

The consequence of homogenization is that the 18S–28S gene usually *behaves* more or less as a SCL.[60,61] even though its copy number is at the opposite end of the scale. There are exceptions, however, in which divergent paralogs, even expressed paralogs, have persisted through speciation, including in some Cactaceae (Harpke and Peterson, 2006, 2007, 2008; see also Garcia et al., 2017).[62] But even perfect homogenization might conceal rather than reveal instances of HGT. This is because one of the two parental forms, via VGT, would tend to convert the other. Nonetheless, the 18S–28S gene (or its homologs) probably is the single most widely exploited locus in molecular phylogenetics across the Tree of Life and across all divergence levels from the most ancient to the most recent.[63]

There seems to be no reason to believe that gene family paralogs should be less susceptible to duplication/extinction (birth/death) than SCLs. This in itself renders likely the same possibility that *each* of the inferred ortholog clusters is erroneous. But on top of this, the duplicity itself should facilitate recombination and/or conversion, which would further mislead orthology inference. Presumably, such processes would occur with *increasing* frequency relative to copy number, as demonstrated by the extreme case of nrDNA.

Because nominal gene family copy numbers are much less than in the case of nrDNA, homogenization ought to be, and evidently is, incomplete. This is because recombination/conversion evidently is a function of copy number, as well as chromosomal distribution/linkage (viz., these processes are consequent to chromosomal pairing). Still, it is reasonable to believe that (partial)

---

[58] The cistron consists of, in order, the 5'ETS–18S–ITS1–5.8S–ITS2–28S–3'ETS genes, where ETS is the external transcribed sequence. The cistrons are separated by the intergenic spacer (IGS).

[59] I ignore here the 5S gene. For a variety of reasons, it generally is not a suitable as a phylogenetic marker (Hershkovitz et al., 1999). But early in the development of molecular phylogenetics (1970s–1980s), 5S-based organismal phylogenies were published in top journals. This owed partially to reductionist "superstition," and partially to the facility with which 5S sequences could be derived using primitive direct RNA sequencing methods. However, the lesson of 5S remains as relevant today as it was decades ago. Because currently analyzed phylogenomic are no better known now than 5S was then.

[60] The apparent homogeneity of nrDNA, including the more variable ITS and ETS regions, may be partially an artifact of PCR homogenization, i.e., each PCR amplification cycle favors the most common fragment, such that forms less common in the genome might not be detected by PCR. It is unlikely that nrDNA gene homogenization is 100%. However, complete genomic nrDNA sequencing is hampered by technical issues associated with highly repetitive DNA (Symonová, 2019; viz. amplification artifacts associated with repetitive DNA and lack of unique signatures for contig assembly).

[61] To underscore this impression, many, if not most molecular phylogenetic studies using nrDNA generally omit reference to "gene family" and "paralogs." Studies that refer to paralogy do not refer to orthology.

[62] In some organisms (I did not find a plant example, but I did not look very hard), the divergent forms express in a developmental- or tissue-specific manner (e.g., Symonová, 2019; other references can be located easily in Google).

[63] As emphasized above, this is not at all to say that nrDNA yields reliable organismal phylogenies or, as analyzed conventionally, even reliable nrDNA phylogenies.

homogenization in gene families proceeds sufficiently to differentially affect paralog evolution at the sequence level[64]…perhaps the worst of all worlds.

Thus, compared to SCL analysis, pseudology is all the *more* likely in phylogenomic analysis. First, there is the inherent difficulty in analyzing orthology (just as for morphological homologues), especially given sparse in vivo gene *functional* analysis across the sample. Secondly, genomic probes recover small fractions of the entire genome, and gene recovery across the samples usually is incomplete. Third, taxon sampling remains a small fraction of that commonly attained using targeted genes. Fourth, as described above, evolutionary dynamics likely are more complex for gene families compared to SCLs. Finally, in practice, it appears that phylogenomic analyses inherently bias in favor of *affirming* an orthological interpretation.[65] Moore et al.'s (2018) criticism of SCLs is a red herring.

Another likely pitfall of phylogenomic analysis is the default idealization of DNA sequence evolution – at any or all loci – as "neutral." This idealization, of course, is inherited from targeted gene phylogenetics. Thus, besides other causes, gene tree incongruence also may be consequent to molecular-level convergence. An example is convergence in base composition, which might cause spurious attraction of sequences in phylogenetic analysis. Site-specific functional convergence is more difficult to detect, especially with inadequate knowledge of molecular-level function. Of course, we are a long ways from understanding all functions of all genes of *any* organism, infinitely less in all organisms in a particular phylogenomic analysis.

However, DNA sequence convergence is both predicted and shown to occur in genes having particular ecological functions in plants sharing similar ecological niches. The point is hardly gratuitous in the case of the Portulacineae data, because Moore et al. (2018) and Goolsby et al. (2018a) enriched their genomic probes with sequences from 19 gene families "known to be important in CAM and C4 photosynthesis." Because Portulacineae vary considerably in their expression of CAM and C4, sequence convergence in these loci might be expected, hence bias phylogenetic results in favor of grouping physiologically/ecologically similar species. Moore et al. (2018) mentioned this phenomenon. I return to the point later in the discussion of taxon relations.

In summary, while the phylogenomics paradigm emphasizes HGT and ILS as the principal, if not exclusive, source of gene tree incongruence, these events actually are "relatively" rare. They are effectively *singular* events that transpired prior to the time of permanent lineage divergence. As noted, at deeper phylogenetic levels, subsequent evolution tends to obliterate evidence of such events. In contrast, recombination, paralog birth-death, gene conversion, TE-mediated evolution, and sequence convergence are *continuous* phenomena. They occur at least somewhere in the genome on the order of between several times per cell division to several times per generation. The cumulative consequence is significant especially in plants, whose indeterminate growth introduces somatic mutations into the germline. This is not to say that HGT and ILS and hybridization have not been any less an important factor in shaping the genomes of modern organisms. The genomes of hypothetical ancestors of modern lineages indeed were likely mosaics comprising genes that were also ancestral to *other* lineages.

---

[64] Thus, the functional differentiation within gene families owes perhaps to *low* copy number, too low to homogenize more completely.

[65] Notwithstanding the limitations of statistical modeling (Hershkovitz, 2021b), it would be useful to model hybridization, gene duplication/loss, recombination, and concerted evolution as parameters in the framework of multi-species multi-locus DNA sequence evolution. Parameter range priors might vary between zero and rates estimated based on existing analyses of combined nrDNA and cpDNA data sets.

The other major source of disagreement is in methodology itself. Phylogenomic methods proposed and applied to parse gene tree incongruencies, including those applied in the case of Portulacineae, remain unsatisfactory, both technically and epistemologically.[66] At least in the case of the nuclear genome, the decoupling between locus (or site) and taxon history, hence the (predicted and proven) inadequacy of concatenated matrix analysis, seems to be broadly appreciated. It is known that discordance even at single loci/sites ("outliers"[67]) can have an inordinate influence in analyses of concatenated sequences (Walker et al., 2018; Neupane et al., 2019) and even in analysis of single locus sequences (Lewis et al., 2016).

Current phylogenomic praxis pursues gene tree congruencies by comparison of concatenated matrix analysis with MSC analysis. The concatenated matrix analyses apply conventional ML reconstruction/bootstrap or Bayesian analysis.[68] The MSC approach (as applied in the Portulacineae analyses; Moore et al., 2018; Wang et al. 2018) emphasizes the proportion of loci that support alternative topologies. In the case where the methods suggest conflicting topologies,[69] theoretical evidence suggests that concatenated matrix analysis is artificial and that MSC is more resistant to outlier artifacts and is to be preferred (e.g., Walker et al., 2018). However, the two approaches as applied in the Portulacineae analyses are not directly comparable statistically.

Their statistical modeling of a nonstatistical phenomenon notwithstanding, the studies of Lewis et al. (2016) and Neupane et al. (2019) are both revealing and insightful with respect to genome and organismal coevolution a propos incongruencies. For example, Lewis et al. (2016) demonstrated the possibility that a MSC tree optimal over all sites/loci might not be optimal according to any locus or combination thereof.[70] This is all the more disconcerting given that the numbers of taxa, loci, and total sequence sites in their simulated and empirical data sets are miniscule compared to real phylogenomic data sets. The sequence evolution model applied likewise is unrealistically simplistic (cf. Hershkovitz, 2021b). Moreover, the only confounding factor they simulated is ILS, not even HGT, much less VGT or convergence.

---

[66] As a disclaimer, for the purposes of the present work, I have not attempted an up-to-date and thorough (or even especially competent) review of phylogenomic analytical theory and methods. After all, the present work is supposed to be Portulacineae, not about phylogenomics. But interpretation of the former necessarily refers to studies of the latter. I comment here very briefly only on the theoretical state of phylogenomics contemporary with the cited Portulacineae analyses, especially because the publications cannot automatically correct themselves to reflect subsequent theory, methods, and criticisms.

[67] "Outliers" in phylogenomic jargon are loci whose cladograms include edges not present in the optimal MSC consensus but whose statistical support strongly biases concatenated matrix analysis in favor of the outlier.

[68] Hershkovitz (2021b) emphasized the established but still unappreciated theoretical inequality between bootstrap proportions (BP) and Bayesian posterior probabilities (PP), as well as between PPs and conventional probabilities (p values), in phylogenetic analysis. I argued that Bayesian phylogenetic analysis is essentially a ML optimization algorithm, such that PPs relate to the probability that a node occurs in the ML tree optimized according to the specified model and parameters.

[69] Here, the term "conflicting" is appropriate because the incongruence is consequent to differential analysis of the same data.

[70] The phenomena analyzed by Lewis et al. (2016) and Neupane et al. (2019 [including Lewis], viz. combined data phylogeny incongruent with all separate data sets) is not novel. The authors merely proposed a solution in the context of their preferred but deeply flawed Bayesian approach (cf. Hershkovitz, 2021b). Chippindale and Weins (1994; cf. Weins and Chippindale, 1994) discussed an analogous case (and analogous solution, viz. weighting) in the context of MP. The method of Neupane et al. (2019) seems conceptually related to that of Chen et al. (2007). Williams and Ebach (2020: 264) believed that the separate/concatenated paradox is an artifact of "tree thinking": "The pooled data…resolve nodes that are *created* (italics theirs)…because of the needs to satisfy the demands of the [method, i.e., in finding the "optimal" tree]…We refer to [the created nodes]…as what they are: *false* [italics theirs], a product of the [method]."

5. Phylogenomics: a tentative diagnosis

In summary, reliability of phylogenomic analyses is undermined by the massive size of the data, and limited comparative knowledge of the analyzed loci. And compared to targeted gene analysis, phylogenomics methods are more "robotic" and data analyses more dependent upon "artificial intelligence." I have not studied evidence for another consideration, reproducibility of both data and analytical results. Targeted gene data and results presumably are more easily reproducible. In any case, the data and analyses are more tractably scrutinized, and the macroevolutionary behavior of the loci generally is better studied across a wide range of taxa.[71]

The nuances of phylogenomic analysis likely are not adequately appreciated except by relatively few specialists.[72] Fewer than the evidently few that adequately appreciate the nuances of targeted gene analysis (cf. Hershkovitz, 2021b). For the rest of the systematics (more so for nonsystematics) communities, indoctrination in phylogenetics dogma ("tree-thinking") seems to have conditioned a priori the credibility of any published cladogram without critical scrutiny or understanding of their technical, much less epistemological, bases.

This is not at all to say that phylogenomic studies are less useful than targeted gene analysis. To the contrary, they are extremely revealing when interpreted in a proper theoretical context. For example, a theoretically predictable and confirmed consequence of whole genome dynamics is that individual (orthologous or pseudologous) locus cladograms (correctly or incorrectly) manifest incongruencies with respect to a fixed organismal cladogram. Viewing genetic phylogenies as "cloudograms" rather than cladograms (e.g., Copetti et al., 2017) is not only more accurate, it provides insight into the ontological nature of organisms and their nonlinear evolution. And even in the prevalent reductionist framework, phylogenomic data yield insights into developmental and physiological genetics at the genome scale, as well as data for advancing the study of molecular evolution.

6. So are phylogenetic conclusions "right" or "wrong"?

The impetus of this work (allegedly) is to evaluate incongruent/inconclusive resolutions of phylogenetic relations among the principal clades of Portulacineae. But the preceding sections might seem to suggest that phylogenetic reconstruction is a pipe dream.[73] In particular, the discussion emphasizes that phylogeny is not an observation (cf. Williams and Ebach, 2020), but an inference, and that, per PEI, it cannot be reconstructed – or "retroactively predicted." This conclusion, in turn, would seem to be at odds with both my historical work and the rest of this work, in which phylogenetic conclusions are considered highly credible if not true. And it is at odds with the notion that phylogenetic history, if unobservable, nonetheless is materially tangible and therefore ought to be *amenable* to scientific discovery (Hershkovitz, 2019b, 2021b).[74]

---

[71] This is not at all to say that targeted gene analysis is correspondingly rigorous *in practice*. This is demonstrated by the sequence alignment of Smith et al. (2017, 2018) discussed above. But scrutiny at least is feasible with relatively simple computational tools. Even so, as I emphasized in Hershkovitz (2021b), targeted gene analysis has been constrained by epistemological more than technical considerations.

[72] I refer in a later footnote to the phylogenomic study of Chen et al. (2015).

[73] Or perhaps a PAUP dream.

[74] I draw analogy here to the field of astronomy, in which heavenly bodies *technically* are not observed directly, but merely inferred, via astrophysical models, on the basis of energy they emitted eons ago. Yet this technicality per se generally does not provoke controversy regarding the existence of these bodies.

However, the preceding discussion does *not* argue that phylogeny cannot be reconstructed. It argues that contemporary evolutionary biology misinterprets biological ontology. This, in turn, has resulted in an erroneous idealization of the evolutionary process, which has been codified into erroneous notions of evolutionary reconstruction. But this is not to say whether or not an evolutionary reconstruction is correct or incorrect.

The purpose of this section is to address this paradox, viz. that theory predicts that evolutionary history is both knowable and unknowable. The first subsection summarizes some critical points relating to the inherently metaphysical field of systematics, in particular to divergent notions in cladistic and phylogenetic theory. This is followed by suggestions that these divergent notions reflect, like all cognition, constitutive cognitive algorithms, or tendencies to perceive the same objects in different but nevertheless somewhat predictable ways. This offers a partial explanation for the observed diversity of notions. But I also suggest, perhaps more radically, that the divergent notions, by virtue of their neurobiological predictability, actually yield ontological insights that are independent of the actual data. I then consider whether the divergent notions themselves predict a phylogenetic ontology that is consistent with that predicted by PEI. In any case, I consider how cognition, independent of data, may limit what we can "know" about phylogeny and what we cannot.

***Inherent dissonance among systematics epistemologies*.** Systematics epistemological diversity by itself accounts for considerable discrepancies in systematics conclusions. Thus, the question of accuracy of systematics conclusions at the analytical level is somewhat of a red herring. Because, however adhered to in practice, the mere persistence of epistemological differences indicates that the "one true epistemology" is not known. There is a tendency, perhaps in all of science, but especially in the more metaphysical fields of systematics, ecology, and evolutionary biology, to afford credibility to conclusions without adequate reference as to how the conclusions were derived, technically, methodologically, or, as importantly, epistemologically. Sometimes conflicting conclusions reflect comparison of apples and oranges.

Over the past two decades, the empirical derivation of cladograms owes to phylogenetic analysis of DNA sequences derived from targeted gene and phylogenomic data. The predominant phylogenetic method applies ML, both conventional and Bayesian. The MSC method in phylogenomic analyses is ML-based. Otherwise, ML is followed by (usually unweighted) MP by itself or in combination with ML. Less commonly applied now are (statistical) distance methods, these usually incorporating ML base substitution models. All of the methods analyze 2-D matrices of DNA sequences aligned using statistical criteria with or without subsequent subjective manual adjustment.

Myriad publications pontificate on the vice and virtues not only among the popular phylogenetics methods, but among various parameterizations and techniques applicable to each. Much of this discussion, however, relates to the computational qualities more or indeed rather than the theoretical bases. The more profound epistemological bases are discussed in relatively few publications that few empirical researchers read or are even capable of comprehending. The "true" phylogeny is the one that the program spits out (cf. Swofford et al., 1996). Different methods often spit out highly similar, but in practice usually not identical phylogenetic trees. Thus, it is impossible to answer the question of "right or wrong" except in reference to the epistemology.

My recent work (e.g., Hershkovitz, 2019a, 2021b) overlooked another school of cladistic philosophy, summarized by Williams and Ebach (2020). This school developed in the late 1970s, but came to be ignored or rejected by the currently predominant schools that publish cladograms. I

elaborate here on this school, because it sheds revealing light on phylogenetic incongruency, resolvability, and, especially, the phylogenetic process.

Williams and Ebach (2020) argued that "true" cladistics pertains to the discipline of taxonomy and classification and not to phylogenetics.[75] They argue that it is agnostic towards the cause of biodiversity (e.g., phylogeny) and is merely a continuation of "natural system" classification established by Linnaeus and other 18th Century (if not earlier) taxonomists. In particular, they argue that classifications themselves, as *hierarchical* arrangements of taxa, are and always have been essentially "cladograms."[76] The cladistic structure is inherent in the classification, regardless of the criterion used to construct it and/or its "accuracy." Willi Hennig (eventually) popularized (but did not invent[77]) the modern classificatory *criterion* of cladistics, viz. synapomorphy, as distinct from mere similarity.[78]

Williams and Ebach (2020) characterized cladistics as the inference of "natural relationships" (see below) as a function of "three-item analysis" whose solution yields "three-taxon statements," the approach abbreviated as "3TS." Thus all taxa of interest are analyzed, but three at a time and considering only the characters informative for the triplet relationship. Thus, for taxa A, B, and C, the possible cladistic relationships are A(BC), (AB)C, and AC(B). The correct 3TS solution groups the pair of taxa sharing the largest number of synapomorphies. Williams and Ebach (2020) demonstrated that the 3TS solution is *inherently* different from (even when coincidentally topologically isomorphic with) that derived from conventional MP analysis of more inclusive taxon sets.[79,80]

---

[75] Confusion here proves to be understandable, since the term "cladistics" popularly is synonymized with "phylogenetic systematics" of Willi Hennig (e.g., https://en.wikipedia.org/wiki/Cladistics#History). Hennig's method is associated with the modern MP school, incarnated in the Willi Hennig Society and their flagship publication, "Cladistics." This school denominates Williams and Ebach's (2020) approach as "transformed cladistics" (e.g., https://en.wikipedia.org/wiki/Transformed_cladistics#Modern_proponents). I cannot locate this term applied or explained in Williams and Ebach (2020); possibly it is referenced indirectly in bibliographic citation. Additional confusion arises from Williams and Ebach's (2020) use of the term monophyly, in their sense purely tautological (i.e., taxa are monophyletic because they are taxa (~ "phyla") and not meant to imply descendency from a common ancestor. William and Ebach's (2020) lexicon reminds me of the contrasting definitions of "democracy." The term has been used to describe popular governance via free elections, or, alternatively, it is a tautological synonym of communism (as in "German Democratic Republic," among others). The same with the term "popular," in the sense of "the people" (as in the "Democratic *Popular*" republics of Korea and Cambodia under Pol Pot) and not in the sense of "popularity." Correspondingly, governance via *popularity* in free elections is referred to by communists (e.g., here in Chile) tautologically as "dictatorship" [of majorities/pluralities]. Excepting when communists win pluralities, as in the case of Salvador Allende. Only *then* does "popularity" suddenly legitimize itself as the incontrovertible arbiter of "democracy."

[76] Accordingly, Linnaeus, De Candolle, etc., were "cladists."

[77] See, e.g., https://en.wikipedia.org/wiki/Cladistics#History; this history is not critical to the present work.

[78] Thus the distinction between synapomorphy and symplesiomorphy and the recognition that the latter are uninformative. Classical taxonomy as a matter of practice often diagnosed taxa on the basis of symplesiomorphy.

[79] The discrepancy owes to several factors: (1) 3TS uses all data relevant to the three taxa; generally these data cannot fit properly into a MP multi-taxon 2-D data matrix without introducing "missing data," which yields MP analysis artifacts; (2) character states always are polarized for 3TS but usually not for MP; (3) apparent character state "reversals" are effectively discarded in 3TS, whereas they take on a life of their own in MP (except in "Dollo" MP, but here the calculation is not the same as for 3TS). Compared to MP, 3TS resembles a multi-level rather than planar chess game.

[80] Williams and Ebach (2020) also noted similarities between 3TS and characteristics of the method of compatibility (which also lacks "homoplasy," a term the authors consider logically redundant with "analogy, as opposed to homology/synapomorphy). The shortcoming of compatibility seems to be the same as MP, viz. use of a *single* 2-D data matrix representing all characters for all taxa, whereas 3TS analyzes taxa three at a time and only for characters informative for those triplets.

Williams and Ebach (2020) argued that what is now called "cladistics" (viz. MP) is not rooted in the science of taxonomy/classification, but rather in both numerical phenetics and pre-Hennig phylogenetic systematics (from Darwin to, e.g., Ernst Mayr and Arthur Cronquist).[81] The latter indeed often parsed classification according to one or more stated or unstated notions of "evolutionary principles." The post-Hennig MP paradigm accepts the notion or at least the ideal of synapomorphy (as opposed to symplesiomorphy) as a marker of cladistic relationship. But, as Williams and Ebach (2020) noted, MP infers synapomorphy secondarily as a function of *ancestry* according to shortest phylogenetic trees[82] derived from numerical analysis of data matrices simultaneously representing all taxa and character data. Thus, synapomorphy follows not from original interpretation of data, but reinterpretation or, in Williams and Ebach's (2020) terminology, "distortion" of the data.

Williams and Ebach (2020) in no way negated the legitimacy and value of the study of phylogeny. But they argued that phylogenetic reconstruction necessarily follows from classification and not, as in current praxis, the reverse. They also argued that current phylogenetic methods derive not from (their concept of) cladistics, but from phenetics. In particular, they argued that MP *technically* is "weighted phenetics," viz. "overall similarity of synapomorphy" determined methodologically via "quantitative phyletics." Analogous terms could be applied to ML. Instead of this phenetic approach, Williams and Ebach (2020) suggested a phylogenetic reconstruction approach that they considered to be free from assumptions and artifacts, and based only on empirical data, viz. development ("time"), morphology ("form"), and geographic distribution ("space").

Williams and Ebach (2020) expressed many viewpoints that are consilient with those I have expressed independently, though not necessarily for the same and sometimes even for contradictory reasons. For example, my criticism (Hershkovitz, 2019a, b; 2021b; see also above) stressed the inadequacy of the tree metaphor to capture organismal phylogeny. Williams and Ebach (2020) professed agnosticism towards the "shape" of phylogeny but, nonetheless, their philosophy of taxonomic science welds them to a tree-like concept, viz. hierarchical classification.

I also agree with Williams and Ebach (2020) to the effect that phylogenetic reconstruction and/or cladistics is an endeavor of logic and reason rather than of mathematics or statistics, and that no mathematical/statistical value inherently reflects phylogenetic truth. Here my opinion follows from my diagnosis of the evolutionary process as idiosyncratic, which Williams and Ebach (2020) might classify as "science fiction."[83] I also agree in part with Williams and Ebach's (2020) view that the mainstream paradigm derives taxonomic conclusions only from dubiously justified mathematical abstractions of thereafter discarded empirical data. However, I do maintain that mathematical abstractions have heuristic value when the amount of data greatly exceeds that which can be parsed

[81] Accordingly, Williams and Ebach (2020) would view criticisms of "cladistics" by Cronquist (1987) and Mayr and Ashlock (1991) as red herrings, because they confound cladistics and phylogenetics.

[82] In practice, phylogenetic analyses generate unrooted trees that are then rooted using outgroup or, less often, midpoint rooting. The rooting then is supposed to specify both the relationships among the taxa and the set of their possible synapomorphies. Williams and Ebach (2020: 98–99) maintained that: "The…unrooted tree approach to understanding *relationships* [italics theirs) amongst organisms is simply misapplied mathematics, as is the graph theory derived from it, when used for systematics/taxonomy…"

[83] PEI is an explanation for the pervasiveness of incongruence (viz. noise or lack of consistent pattern) in phylogenetic evidence, i.e., it proposes that "noise" is an *informative* part of the signal of evolution (Hershkovitz, 2019b). Williams and Ebach (2020: 365) wrote, "Finding meaning (e.g., a process) when there are no patterns (i.e., no evidence or 'noise') is bordering on what can only be described as science fiction." But PEI is not fantasy. It emerges as a logical and *unavoidable* consequence of biological (autopoiesis), physical (hierarchy), and mathematical (chaos) theory. Williams and Ebach (2020) demonstrated no qualification for adjudicating such theories.

purely by inspection. Nowadays, this is essentially always. Abstractions can identify unappreciated aspects of the data.

From a theoretical standpoint, Williams and Ebach's (2020) complaint about theoretical assumptions is somewhat of a red herring. This can be appreciated by working backwards from the currently favored reductionist approach, ML. ML (and its poor cousin, ME) presumes that characters or classes thereof evolve according to a common and constant stochastic process. Character subsets are reduced into as few classes (parameters) as can be discriminated statistically.[84] Unweighted MP effectively presumes that no common uniform process or mechanism links evolution among the various characters (Huelsenbeck et al., 2008).[85] In this case, the number of parameters equals the number of characters. So what of 3TS? It goes further than MP in presuming that *relationships* among *organisms* ("the *cladogram* of life") are not determined by any common uniform process or mechanism.[86] The number of parameters is a function of *both* the number of *taxa* and the number of characters. Thus, the number of parameters might increase disproportionately relative to the linear increase in taxa/characters and approach "infinity" more rapidly than for MP.[87] There is a continuum here, though not a linear one, from fewer to more parameters.

In fact, hyper-parameterization of 3TS precisely explains its computational limitations described by Williams and Ebach (2020). Reduction of data for all taxa into a single 2-D matrix vastly simplifies heuristic optimization[88] of MP and ML (and ME) methods. 3TS[89] is far more computationally intensive, and it may not be ever feasible for the numbers of taxa and characters currently analyzed using single matrix approaches (cf. Williams and Ebach, 2020). I argue here that this is because each 3TS is a parameter.[90] Extension of 3TS to DNA data evidently also is (otherwise) problematic. Williams and Ebach (2020) acknowledged this problem but offered no resolution.[91]

---

[84] And in the stochastic model case, these generally are misspecified technically, while in the PEI case, they are misspecified epistemologically.

[85] Mainstream dogma maintains that the "no common mechanism" assumption renders MP is "statistically inconsistent" and therefore inherently unreliable in phylogenetic reconstruction. This also is a red herring, because no commonality of mechanism, much less the "true" one, ever has been demonstrated, except as a statistical epiphenomenon. Thus, ML "fake models" *ultimately* are no less inconsistent than MP.

[86] When transposed to the context of phylogenetics, this observation corresponds to my observation that, per PEI, evolution is an infinite-order Markov process and not, as predicated by mainstream molecular evolutionary dogma, a first-order Markov process.

[87] I have not derived the actual relationship, though it can be concluded that the increase in parameters (number of 3TS) is not linear with the linear increase in taxa/characters. Not coincidently, the relationship is *idiosyncratic*.

[88] Branch-and-bound and exhaustive searches for large numbers of taxa remains computationally prohibitive. Also, ML and ME programs are feasible only because they apply additional mathematical shortcuts and approximations.

[89] Again, 3TS predicates to infer not phylogeny of, but of *relationships among* (~classification) taxa, the latter then establishing the foundation for inferring the former. Per Williams and Ebach's (2020) arguments, mainstream phylogenetics praxis is effectively self-defeating because it short-circuits analysis of relationships. Williams and Ebach (2020: 99) made this point in response to a comment by Joe Felsenstein: "Oddly, by viewing systematic data as relational [i.e., for classificatory purposes *only*] instead of linear it becomes both 'less ambitious' [per Felsenstein], by not attempting to recreate, reconstruct or 'infer' phylogeny, and at the same time 'more powerful' [per Felsenstein, in the sense of evaluating *all possible* trees], by providing an exact method for biological classification."

[90] Williams and Ebach (2020) made this point indirectly by noting the extraordinary numbers of rooted trees for any number of unrooted.

[91] This is far beyond the scope of the present work. The dilemma may root in the ontology of base substitution and its interpretation in terms of homology. In particular, base substitutions technically are not modifications of prior states, but replacements. The prior state indeed is *lost* as the new one is gained, and the process is indeed

But moving on from this point, Williams and Ebach (2020) claimed that cladistics (viz. classification) must be based only on qualitative structural similarities. They rejected quantitative similarities, such as size, stating, for example, that "all leaves change size during growth" (Williams and Ebach, 2020: 185). As this change (a process) is not evident in the specimens being classified, this "principle" amounts not to an empirical observation, but an assumption or a dogma – the very sort of criteria Williams and Ebach (2020) *eschew* in cladistic analysis.

Likewise, Williams and Ebach (2020) accepted, even advocated, the use of characters from different developmental stages of the organism. Development is a *process*, albeit observable, but not, in any case, itself a structural character. The relation of developmental stages thus must be *inferred* and based on *assumptions*. Why, then, Williams and Ebach (2020) reject inferences based on *other* logically inferred processes, such as HGT or phylogeny…*or autopoiesis*… is not clear, though I attempt to explain it below. *Strict* adherence to Williams and Ebach's (2020) criteria might well result in "clades" comprising, e.g., juveniles or males versus adults or females. In other words, the correct classification of these different structural stages/forms requires not merely observational, but also *logical* inference.

I dwell here on the viewpoints of Williams and Ebach (2020) partly for the property described above, as well as observations described in the next section. Many of Williams and Ebach's (2020) talking points are not only valid, they corroborate PEI. The most notable among these is the empirically evident but generally ignored expansion/contraction of character space.[92] This not only explains, it predicts the observation of, for example, and as emphasized by Williams and Ebach (2020), the partial to mutual exclusivity of morphological characters used to classify and/or reconstruct phylogeny of different groups of organisms. I argued above that the same applies to DNA data. Evolutionary patterns are different in different parts of the tree, because context also differs. This is a prediction of PEI. It is not science fiction.

The upshot of this section is that, assuming phylogeny, the accuracy of phylogenetic analyses cannot be evaluated purely in terms of the results of analyses of empirical data and/or the method and/or the model. There remain fundamental epistemological differences. As long as these remain unresolved, the best that can be hoped for is in terms of reconciliation, as discussed below.

Dissonance in systematics…..is it a "brain" thing? The discussion thus far might seem to have deviated significantly from the titular purpose of this work, viz. reconciliation of disagreements on the backbone topology among published cladograms/phylograms of Portulacineae. This is in order to elaborate on the general causes of dissonance before applying this knowledge to the specific instance. I have considered already dissonance owing to incongruence among characters, then at another level methodological disagreements, and finally epistemological disagreements. These already challenge the very notion of phylogenetic resolvability.

---

reversible. But Williams and Ebach (2020) may be correct in their assessment that the process itself is irrelevant and that the diagnosis of homology can be based only upon relationship.

[92] I call attention here to the phylogenomic study of Chen et al. (2015). These authors advocated a phylogenetic question-specific, rather than global, approach to phylogenomic analysis. In particular, they argued that the global approach, analysis of an all-gene X all taxa 2-D matrix, likely introduces noise/error at particular phylogenetic nodes of interest. The conclusions not only document phylogenetic character space expansion/contraction, they even less wittingly validate the arguments of Williams and Ebach (2015) and the 3TS approach.

The present section discusses another cause: dissonance owing to neurobiological factors. My discussion is speculative and not corroborated here formally with references. But it is axiomatic and uncontroversial that notions of reality must be influenced by the physiology of cognition, perhaps even such notions as "real numbers."[93] Here I speculate on how constitutive cognitive algorithms might affect notions in systematics in a way that generates dissonance. And I propose that cognition-based dissonance itself may shed light on the nature of phylogeny.

I begin by considering the notions of classification emphasized by Williams and Ebach (2020). To summarize, they distinguished their concept of natural classification, which they said must be *discovered* (see also below) from artificial classification, which is manmade and *imposed* for anthropocentric purposes. They argued that this natural classification, by virtue of its inherent hierarchy, itself represents the natural or perhaps "true" cladogram of life. And they argued that the natural classification is prerequisite for interpretation of the *causa ultima* of biodiversity, be it phylogeny or anything else. They concluded, thus, that a hypothetical phylogeny can be derived from the natural classification, but not vice versa.

I suggest that classification itself is a constitutive cognitive process that, while essential to the knowledge system, is distinct from cognitive processes underlying scientific reasoning. I conjecture that classification is essentially a cognitive reflex, that we cannot avoid it, and that it is *inherently* hierarchical. Thus, we can classify things differently, but each classification is constrained by hierarchy, because the mind cannot deal with a nonlinear classification.

In this sense, Williams and Ebach's (2020) arguments become a red herring. The "naturalness" they idealized is "natural" to the *mind* and not (necessarily) to the Order of Nature. It is *cognitively* natural. Williams and Ebach's (2020) classification is, in fact, *just as* manmade as and not qualitatively different from a so-called artificial classification. And it is indeed for anthropocentric purposes, the purpose of cataloging life in a way that proves to be most *utilitarian* in terms of stability and predictability. Otherwise, why bother?

However, the denomination "natural" *necessarily* specifies its attribution to "nature." Thus, a truly natural classification, whether or not it is the "true" one, and indeed if a true classification even is possible (see below), necessarily implicates *some* model of nature, such as phylogeny, even if "no model" is presumed. Interestingly, Williams and Ebach (2020) did not explain adequately *why* "natural classification" pertains only to living beings and not to nonliving beings. In other words, they did not explain the *justification* for the very science they advocate.

The *anthropogenic* origin of biological taxonomy evidently was recognized in ancient philosophy, as in Genesis 2: 19, "And out of the ground the Lord God formed every beast of the field, and every fowl of the air; and brought them unto Adam to see what *he* [Adam!] would call them: and whatsoever *Adam* called every living creature, that was the name thereof (King James Version; italics mine).[94] This conveys the belief that God created the organisms, but man created the classification for the purposes of mankind and not for God. In other words, there is no such thing as a truly natural classification of life, because, in this context, this only could have been a classification authored by life´s supposed creator, God.

[93] There is more than ample literature on the neurobiological/philosophical interstice, but this is far beyond the scope of the present work. I leave it to the reader to find in Wikipedia. But it is not "science fiction" as Williams and Ebach (2020) might have it to be.

[94] The original authority here is not known, because this book documents beliefs that, before its writing, had been propagated intergenerationally orally.

Alternatively, such a classification could be *inferred* by *discovery*, as Williams and Ebach (2020) suggested (see above). In other words, without knowledge of the model, the Order of Nature might be inferred through the process of classification. But one cannot escape the requisite *assumption* that nature is indeed ordered, and, in the case of Williams and Ebach (2020), ordered in a bifurcate hierarchy. Otherwise, what is it that Williams and Ebach (2020) predicate to be *discovering*?

But if cognition constrains classification to be hierarchical, the red herring emerges, and classification itself is not so helpful. Because then indeed it is constrained to the form of a cladogram. As Williams and Ebach (2020) emphasized, the cladogram is specified by the classification itself. The lines, which take on a tree-like geometry, would be the brain´s simplest way of graphic summary. Indeed, Williams and Ebach (2020) emphasized that a cladogram, unlike a phylogram, is *not* a tree, effectively because the branches of a cladogram are dimensionless abstractions that do not correspond to any *physical* connectivity/continuity specified by a tree. In contrast, in a phylogram, the branches correspond to *physical* genealogical history, which cladograms do not portend to portray. This explanation for what a cladogram is not, as opposed to what it is, is consistent with the notion that the branches of a cladogram represent a subliminal cognitive device.

Williams and Ebach (2020) might (or might not) be philosophically correct in their claim that a classification/cladogram is theory-free. But it is clear that classifications *canalize* the conception and development of theories (Leonelli, 2013). Williams and Ebach (2020) argued that phylogenetics must base upon classification, never the reverse. They also highlighted Darwin's (1859) observation that his notion of phylogeny, viz. as a tree, already was anticipated in Linnaean classification. Thus, classification itself is the source of the *idealization* of the tree form in modern phylogenetics and thus also the obsession with bifurcate tree *resolution*. Williams and Ebach (2020) claimed agnosticism towards whether phylogeny is tree-like. Yet, if classifications are cognitively-constrained to be tree-like, they inherently constrain the tree-like interpretation of phylogeny conceived by Darwin and pursued obsessively in modern phylogenetics.

There is an irony here. Williams and Ebach (2020) argued that classification effectively is the "virtue" and mainstream phylogenetic analysis the "vice" of systematics. They recognized myriad empirical evidence that phylogeny, or rather the collective evolution of characters used to resolve phylogeny, is not precisely tree-like.[95] But they dismissed the significance of this evidence, as well as that of studies that attempt to represent this evidence graphically. They argued that the evidence represents nothing more than characters incongruent with a classification. To them, apparently, incongruent characters are of no use to the advancement of systematics as a science. However, the advancement of *science* pertains to the area of cognition associated with the application of constitutive algorithms of logic and reason towards the goal of optimizing useful *knowledge*. In *this* cognitive realm, incongruent characters become informative. Williams and Ebach (2020) thus seem to reject logic and reason (~ science) in favor of a sort of "me Tarzan, you Jane" cognitive classificatory reflex.

Neurobiology may provide a clue to explain dissonance generally, viz. different interpretations of the same phenomenon at any level. Thus, phylogenetic conclusions may differ in a way that may be specific to the analyst, to the method, among methodologies, and among epistemologies. And the same phenomenon, dissonance *itself*, might originate from the perspective of epistemology and then

[95] For example, and unlike phylogeny (generally speaking), and even classification for that matter, HGT is an empirically observable and reproducible phenomenon. Its inference requires no more and probably less of a leap of faith than the diagnosis of homologues so fundamental to both cladistic and phylogenetic analysis. Williams and Ebach (2020) seemed to relegate HGT to the realm of ad hoc assumption. In the present context, this seems to owe to their cognitive reflex to conserve their perception of the natural order as hierarchical.

percolate down to the level of the empirical analysis. In either case, the dissonance creates a sense of uncertainty/ambiguity.

However, it is known in neurobiology that dissonance indeed emerges in the case where the phenomenon is truly irresolvable. Among the many studied examples is the famous Necker Cube (Fig. 2). This is simply a planar drawing that the viewer perceives as three-dimensional. This also owes to constitutive cognitive algorithmics. However, the Necker Cube yields two different three-dimensional resolutions. Depending upon the viewer, usually one or the other is perceived first, then the alternate resolution usually can be perceived by cognitive concentration. The point is that the drawing in the planar sense is *unambiguously unresolved*. The apparent *ambiguous resolutions* are consequent to cognition.

In the context of the present work, interpretation of the apparent ambiguity in resolution of the Portulacineae backbone phylogeny, the Necker Cube raises some interesting questions. Is it possible that the cognitive phenomenon emergent in the Necker Cube and countless simple optical illusions…also emerges in the perception of highly complex phenomena…such as the Order of Nature? Is it possible that dissonance itself might represent simply alternatively biased but incomplete perceptions of that *projected by* a singular object or phenomenon? Can the dissonance itself…as in perception of the Necker Cube…help us better understand the phenomenon independent of our possibly distorted empirical observations?

I believe that the answer to these questions is "yes." I offer here an example, again taken from Williams and Ebach (2020). Figure 3 is taken from de Candolle (1828) and illustrates his concept of "natural relations" among Crassulaceae. From the standpoint of cladistics, the illustration is ambiguous and, in any case, the lines suggest that the family cannot be resolved as monophyletic (see commentary by Nelson and Platnick, 1981). Candolle (1828) classified Crassulaceae in two major groups, "Crassulaceae Legitimae" comprising those members within the apparent plane of the circle, and "Crassulaceae Anomale" comprising the taxa that appear to be "floating" above the others in the upper right quadrant. He subdivided "Crassulaceae Legitimae" into taxa *firstly* separated by the left and right semicircles, and *secondarily*, each of these into taxa separated in the upper and lower halves.

Williams and Ebach (2020) wrote, "…it is reasonably easy to see how Candolle's written classification relates to the diagram." Clearly there is a relationship. But considering just the four quadrants, there emerge not one but *two* equivalent (cladistic) classifications, the *other* emphasizing *firstly* Candolle's (1828) *secondary* division between the top and bottom halves of the circle. And this, aside from other details, indicates that Candolle's (1828) notion of relationship, his classification notwithstanding, was *not* hierarchical. His classification, therefore, was indeed intentionally artificial, an arbitrary key to the taxa. It seems that Williams and Ebach (2020) first *conditioned* their perception on Candolle's (1828) *classification*, and then subconsciously *preferred* to perceive Candolle´s diagram in this way. Even though the diagram itself is cladistically unambiguously unresolved. Necker Cube.

The other major question raises a possibly more radical suggestion. This is the question of whether the ontology of a phenomenon can be derived from dissonance alone, without empirical study of the phenomenon itself. For example, the ontology of the Necker Cube should emerge as the optimal solution to its conflicting perceptions, viz. a cube from one perspective versus a cube from another. Empirical evidence, examination of the cube, biases its perceived ontology, hence perhaps it is better left unobserved. The last point is critical, because normally scientific investigation indeed appeals directly to empirical data. Yet different investigators often– perhaps axiomatically always – reach different conclusions.

The reason I raise this question is because I have argued that PEI renders phylogeny inherently ambiguous. This owes to the chaotic component of evolution, and the fact that chaotic functions cannot be reconstructed exactly unless the exact function *and starting conditions* are known. However, cognition is limited constitutively to algorithms that necessarily yield or at least converge upon exact solutions, e.g., statistics or cladistics sensu Williams and Ebach (2020). This is understandable. Ambiguity leads to indecisiveness, which leads to cognitive dysfunction. But as I have argued, such approaches to *phylogenetics* are inadequate and bound to yield wrong solutions. I suggest here that comparison of the dissonant conclusions from dissonant approaches, alone and without reference to empirical data, is sufficient to reach the conclusion that phylogeny inherently is ambiguous, as predicted by PEI. At the same time, it projects various (but illusory) patterns that different minds resolve variously (and forcefully) as real.

Phylogenetic reconstruction: not resolution, but reconciliation. Here I return to the paradox that phylogenetic history is materially tangible, hence discoverable, but at the same time, theoretically irresolvable. As for material tangibility, it can be ascertained that material evidence exists, sometimes in the form of preserved or permineralized ancient forms of life, but mostly in the phenotypes and genotypes of present day organisms. Genetic and epigenetic inheritance of phenotypes/genotypes remains a fundamental biological lemma in phylogenetic reconstruction. But irresolvability manifests both theoretically (see above and Hershkovitz, 2021b) and empirically (i.e. by myriad phylogenetic relations that continue to resist resolution). How can these seemingly contradictory observations be reconciled? That is exactly how: an approach, hardly radical, that emphasizes *reconciliation* rather than resolution.

Reconciliation is not the same as resolution. It is equivalent to the consilience approach of Wilson (1998). In the present case, I suggest the notion of reconciliation as an approach to reconcile the evidence *whether or not* that evidence resolves cladistically. This deliberately calls attention to and compensates for the manifest cognitive preference for resolution. Parts of a phylogenetic history can be reconciled as not resolvable. Just as for resolvable nodes, an explanation can be sought in the evidence. Alternative resolutions derived in individual analyses can be reconciled first in terms of analytical artifacts. To the degree that artifacts can be discounted, the Necker Cube effect emerges as an explanation. Which is to say that *resolution* derived from individual analyses may be illusory.

At the same time, diverse epistemologies, methods, and data often derive some of the same clades. In this case, it seems reasonable to accept conclusions that, in diverse empirical analyses, are the least falsified. This may go against the grain of methodological "purists".[96] But in view of the theoretical discussion here, it emerges as a most logical and reasonable approach. While PEI purports that the entire phylogenetic history cannot be recovered, this does not mean that some clades, effectively "fractals," cannot be recognized. Different epistemologies, methods, and data effectively represent biased filters with which phylogeny is discerned. It seems reasonable to afford ontological reality to objects that appear regardless of bias, viz. filters that otherwise might render the object invisible.[97]

[96] In my four decades of studying the systematics literature, I recall coming across opinions that reject this "pluralistic" approach. In other words, no significance is afforded to the coincidence of cladistic/phylogenetic conclusions derived by different approaches. Some workers maintain that a particular approach is correct, and summarily dismiss the value of other approaches even if they yield isomorphic results. I cannot here recall specific examples, though Williams and Ebach (2020) seem inclined towards this view. This tendency itself may reflect the mind's inability to deal with multiple truths.

[97] Of course, it is possible that within certain parameter space, different perspectives are bound to coincide. This is another factor to investigate and reconcile.

Hershkovitz (2021b) emphasized that the truth of a phylogeny can never be evaluated per se in terms of any quantity or statistical value. Evaluation always is subjective, even when supposedly objective criteria are applied. All objectivity ultimately rests upon subjectivity. Given PEI, objectivity in phylogenetic analysis inherently is misleading. Again, a lemma of PEI is that, strictly speaking, biological organisms are not discrete *objects*. Organisms are *stages* in the dynamic life process. They cannot be analyzed in the same way as inanimate *objects* whose fate is determined purely by predictable laws of physics.

I suggest that *both* resolvability and irresolvability represent ideals in phylogenetic reconstruction, because both are predicted. For example, clearly the "anomalous zone" (see above) represents the norm of diversification at the population/species interface. Here, genes cannot resolve phylogeny, because multiple lineages diverge over short evolutionary time. Also, it must be appreciated that, during evolution, phylogenetic evidence does not accrue like rings on a tree. To the contrary, it *sublimes*. This is a thermodynamic process. This view might seem pessimistic from the perspective of modern phylogenetic dogmatists, but it is not pessimistic at all. However good or bad, phylogenetic reconstructions must be appreciated when they are resolved cladistically and when they are not. Both cases represent "resolutions."

It must be appreciated that the *logic* of mathematical/statistical approaches is incompatible with ontology of biological organisms. In my view, this incompatibility *emerges* in empirical analyses when supposedly better models fail to yield better results. Better-fitting models sometimes yield results more erroneous than poor-fitting models (Adabi et al., 2019) and, in the case of Bayesian estimation, optimized models can yield spurious results (Alfaro and Holder, 2006; Grünwald and van Ommen, 2017; Autzen, 2018; Yang and Zhu, 2018). The explanation of this paradox is that statistical models necessarily oversimplify evolution in exchange for statistical power. "Better" models are not "true" models, but simply models that resist rejection for lack of data. They do not perform necessarily better than rejected "worse" models, because in the bigger picture, both are comparably bad.

But none of this is to say that the mainstream mathematical/statistical approaches have no heuristic value. Rather, it is to say that, given PEI, phylogenetic truth never can be adjudicated by mathematical/statistical analysis. Still, the various mathematical/statistical values that are generated in phylogenetic analysis can be useful. This is because, as noted, the amount of data in current analyses is too large to visualize mentally. But the utility of the values emerges only in the context of reconciliation. Firstly, the theoretical basis of the values must be understood. This is no small challenge, as evident in Hershkovitz (2021b). Then the analysis can be undertaken with various modifications of data and parameters to see if the values are stable or if they change substantially. In the latter case, the "problematic" data can be identified and reconciled. Reconciliation is the key.

C. Preliminary reconciliation of evidence for the Portulacineae backbone phylogeny

The preceding notwithstanding, clearly I predicate that some clades within Portulacineae are sufficiently well evidenced materially to be codified taxonomically. This is to say that a name can be given to a group of organisms that are not necessarily descendents of a *single* common ancestor, but nonetheless that can be considered to comprise organisms that have evolved independently of any and all sibling clades. But other aspects of Portulacineae phylogeny, notwithstanding intensive analysis, are not thoroughly or at all resolved cladistically. But in the framework of reconciliation, I do not view this as problematic.

The focus here is on genetic data. Some aspects of Portulacineae phylogeny were resolved on the basis of morphological data (e.g., Hershkovitz, 1993), but most of the established nodes of the phylogeny were not. This might be in part because morphology had not been studied critically. But more often, it seems that relations established using genetic data might not have been established any other way. These include the outgroup relations of Portulacineae and, more broadly, the composition of Caryophyllales and relations of the latter to other angiosperms.[98] Within Portulacineae, currently established relations of genera such as *Calyptrotheca*, *Hectorella*, and *Lenzia* were not anticipated by morphology. It is possible, however, that more detailed anatomical analysis might have helped phylogenetic reconstruction (e.g., Carlquist, 1997).

1. Outgroups of Portulacineae

There appears to be consensus that Molluginaceae (sensu Thulin et al., 2016) are sister to Portulacineae (Moore et al. 2018; Smith et al., 2018; Wang et al., 2018; Yao et al., 2019). These together comprise the "Portullugo clade" (Edwards and Ogburn, 2012). Outgroups of Portullugo include a clade comprising ten families [Nyctaginaceae, Aizoaceae sensu Hernandez-Ledesma et al. (2015), and much of Cronquist's (1980) polyphyletic Phytolaccaceae], and Limeaceae.

2. Phylogeny of Portulacineae

Four regions in Fig. 1 are critical to this discussion. These are these are the relations of: 1. Montiaceae; 2. Halophytaceae, Didiereaceae, and Basellaceae; 3. The ACPT clade; and 4. the basal lineages of Cactaceae.

Montiaceae relations. Hershkovitz (1993) first suggested that "western American Portulacaceae" (i.e., Montiaceae without *Phemeranthus* and *Rumicastrum*) might be sister to remaining Portulacineae. But with the inclusion of data from *Phemeranthus* and *Rumicastrum*, this general idea was corroborated (but without bootstrap support) with ITS sequences (Hershkovitz and Zimmer, 1997). This relation later was confirmed in multiple analyses and now is generally accepted (Moore et al., 2018; Goolsby et al., 2018a, Smith et al., 2018; Wang et al., 2018; Yao et al., 2019).

However, Moore et al. (2018) remarked that "Montiaceae consistently appears as sister to the remaining Portulacineae, though with lower support in both ASTRAL [= MSC] and concatenated analyses than we would have predicted." The BPs for this node in analysis of the various data sets ranged from 74–78 for the MSC analyses and 67–81 for the concatenated analyses (Moore et al., 2017, 2018).

Moore et al. (2018) also reported a low "concordance factor" (CF; 0.28) supporting this basal split of Portulacineae. The CF calculation is supposed to reflect the proportion of loci supporting particular topologies. Unless very high or very low, the absolute value is less important than the value relative to the CF of alternative topologies and to the proportion of loci uninformative for that node.[99] The low CF in this case appears to reflect partially the relatively high CF for the conflicting clade Halophytaceae + Montiaceae (0.19). Thus, ignoring the problematic Halophytaceae yields the standard result. Wang et al. (2018) did not sample *Halophytum*. However, while these authors did not provide a detailed analysis of gene concordance for the basal split of Portulacineae, they illustrated (Fig. 2)

[98] Current understanding is radically different from that of the pre-molecular era. See, e.g., the chapters in Behnke and Mabry (1994).
[99] An analogous relation characterizes BPs (Hershkovitz, 2021b).

considerable gene tree incongruence for the clade Portulacineae-minus-Montiaceae. This will be discussed further below.

As noted above, it had been suggested that Portulacineae-minus-Montiaceae do share a morphological similarity, viz. parallelocytic/anisocytic (developmentally mesogenous) stomata, versus brachyparacytic and other developmentally perigenous forms of Montiaceae (Nyffeler, 2007; Hershkovitz, 2019a). But this is erroneous: *Phemeranthus* has parallelocytic/anisocytic stomata. Also as noted, *Phemeranthus* has other traits typical of Portulacineae-minus-Montiaceae and no clear synapomorphy demonstrating its relation to Montiaceae.

Relations within Montiaceae were discussed by Hershkovitz (2019a), emphasizing the results of Goolsby et al. (2018a, b). Hershkovitz (2019a) recognized two subfamilies, Phemerantheae (tentatively including the poorly known genus *Schreiteria* Carolin) and Montioideae sensu Hershkovitz (2019a), which includes remaining Montiaceae. Montioideae thus is characterized by the mesogenous stomata, as well as more strikingly rosettiform morphology and generally broad leaf bases. Also, these taxa seem to have originated along the west slope of the south-central South American Andes mountains. The ancestral distribution of remaining Portulacineae, including Phemerantheae, appears to have been on the east slope of the Andes. Thus, the evidence suggests that the stomatal form of *Phemeranthus* is the ancestral condition of Portulacineae as a whole.

***Basellaceae, Didiereaceae, Halophytaceae*.** The consensus opinion is that these families comprise the immediate outgroups of the ACPT clade. But here gene trees differ significantly with respect to a fixed organismal phylogeny. All five of Moore et al.'s (2017, 2018) data sets yielded both concatenated and MSC trees showing > 95% support the relation of Didiereaceae as sister to the ACPT clade. Support was less for the sister relation of Basellaceae + Halophytaceae to the ACPT + Didiereaceae clade. Support also was less for the sister relation of Basellaceae and Halophytaceae, ranging from 32–81% for the various concatenated data sets and 78–88% for the various MSC data sets.

Using much of the same sequence data, Goolsby et al. (2018a) found rather lower support for the relations found by Moore et al. (2017, 2018). The MSC analysis yielded 55% BP for Didiereaceae + Cactaceae, 69% BP for Halophytaceae + Basellaceae, and a mere 37% BP for Portulacineae-minus-Montiaceae. However, the authors remarked that "we performed very little curation of this data…. and predict that removal of the most poorly sampled loci would improve support along [Moore et al.'s, 2018] backbone."

The transcriptome analyses (Wang et al., 2018) found that ca. 60% of informative loci supported a constrained sister relation between Didiereaceae and Cactaceae, while ca. 40% supported a sister relation between Didiereaceae and Basellaceae. Considering only gene trees differing by $\Delta ln L > 2$, the data favored the first relation by nearly two to one. Still, this demonstrates a considerable number of gene trees favoring the latter relationship. It should be noted, however, that the transcriptome data evidently included both nuclear and chloroplast loci. The simultaneous age/topology analysis of cpDNA *matK* (Lendel et al., 2013) found Didiereaceae as sister to Basellaceae, whereas their conventional MP and Bayesian reconstruction of the same data did not resolve Basellaceae-Didiereaceae-Halophytaceae relations.

Meanwhile, the concatenated analysis including both nuclear and chloroplast targeted genes (Smith et al., 2018) yielded 71% BP for an unresolved clade comprising Halophytaceae, Basellaceae, and Didiereaceae. However, I discussed above sampling and alignment shortcomings in this work. Analysis of the chloroplast genome sequences (Yao et al., 2019), which did not include

Halophytaceae, yielded 98% BP for the sister relation between Basellaceae and Didiereaceae. This relationship also emerged, though with considerably less support, in Anton et al. (2014, cf. Brockington et al., 2009; cpDNA, nuclear nrDNA loci). Ocampo and Mair-Sánchez (2018; cpDNA) obtained 0.95–1.00 PP for Basellaceae + Didiereaceae, with Halophytaceae unresolved at the base of Portulacineae.

Arakaki et al. (2011; ca. half of the cpDNA genome, nuclear *phyC*) obtained only 45% BP for a clade comprising Halophytaceae, Basellaceae, and Didiereaceae, and only 62% for a clade comprising the latter two. However, they obtained 62% BP for a polyphyletic Didiereaceae, *Calyptrotheca* partitioning with Didiereoideae (classical Didiereaceae), but *Ceraria* and *Portulacaria* with Basellaceae.

The results of Moore et al. (2018) and Goolsby et al. (2018) might have been influenced by a sort of "ascertainment bias." These studies used sequence "baiting" to enrich their genomic DNA sample with sequences from CAM photosynthesis is most strongly expressed in many taxa of Didiereaceae and the ACPT clade, and C4 is unique to the latter, viz. Portulacaceae s. str. (Holtum et al., 2017). Only weakly expressed facultative CAM was detected among Basellaceae only relatively recently (Holtum et al., 2018).[100] Didiereaceae are decidedly more arid-adapted than Basellaceae.

It seems reasonable to speculate that the baited genes might be, via convergence, more similar in taxa sharing more similar physiology and ecology. In fact, Goolsby et al. (2018a) detected such convergence in the CAM/C4-related genes included in the Portulacineae phylogenomic analysis and cited previously discovered convergence evidence. Xerophytic plants converge morphologically and ecologically; there is no reason not to expect corresponding convergence at the DNA sequence level.

This hypothesis might explain the difference between the presumably stronger attraction between Didiereaceae and ACPT baited sequences in Moore et al. (2018) versus the proportionally more functionally neutral transcriptome sequences in Wang et al. (2018). However, molecular convergence might affect loci more broadly. Indeed, Wang et al. (2018) discussed evolution at other loci, especially those involved in abscissic acid regulation, that are thought to relate to xerophytic adaptation.

---

[100] Facultative CAM has been more difficult to detect experimentally, hence may be phylogenetically more common among angiosperms than suggested by the distribution of strongly expressed CAM (Edwards, 2019; Winter, 2019; cf. Gilman et al., 2021). According to Winter (2019), such "intermediacy" between constitutive C3 and CAM expression does not manifest as intermediacy in gene sequence similarity. [As an aside, Edwards (2019: 1744; and earlier works) recognized that statistical evolutionary model parameters, perhaps statistical models generally, are "especially susceptible to *reification* {italics mine}." I suggest that in practice, they are susceptible to *deification*. The point is well taken: the view emphasized here is that the value of models is heuristic. Edwards (2019: 1742) introduced this work with the question, "*Are evolutionary outcomes predictable?*" This work does not provide (or perhaps withheld) the unambiguous answer, which, per PEI (and thus Simon Cowell) is "NO." Edwards (2019: 1743) predicated that analysis of repeated (i.e., "statistically significant") convergence "illustrates….how *natural selection* {italics mine} may strongly favor particular organismal configurations." Per PEI, this reified/deified assertion is DOA ("dead on arrival"; Hershkovitz, 2019b). The evidence Edwards (2019) discussed is more profitably interpreted in the framework of PEI, in particular the discussion of the relation of anatomy to biochemistry, an excellent example of hierarchy theory. The same can be said of Gilman et al.'s (2019) discussion of what they call developmental genetic "modules," whose "*simpler* units…explore mutational space {italics mine}." Gilman et al. (2019 market this work under the heading of "evolvability," a tautology, since that which has evolved *must* have been evolvable. They also market *Portulaca* photosynthesis physiology as a "model system." Since all plants photosynthesize, are any of them *not* "model systems"? But in science, marketing indeed pays off.]

In summary, available genetic analyses do not seem to support a single resolution of the *organismal* relationships of these three families, not to other Portulacineae, nor to each other. The strong plastome data support for Basellaceae + Didiereaceae (Yao et al., 2019) is striking, and no molecular "funny business" was reported. In other words, while plastome sequences recombine infragenomically near the inverted repeat region and much less frequently intergenomically (e.g., Sullivan et al., 2017), there is no evidence that these phenomena have biased the phylogenetic reconstructions cited here.[101] Meanwhile, targeted gene trees by themselves do not seem to strongly resolve Basellaceae and Didiereaceae relations one way or another. While the nuclear genome and transcriptomic loci seem to favor Didiereaceae + ACPT, many factors might bias these data. The conclusions of these studies are not yet convincing.

Floral morphological evidence might suggest that Basellaceae and Didiereaceae are sister taxa. But while it is true that the small size/number of floral organs and small number of seeds are potential synapomorphies of these families, this interpretation depends upon the ancestral condition, which cannot be considered known. Larger sizes/numbers would appear to be ancestral in the adjacent Montiaceae and ACPT clades. However, the smallish perianth might be symplesiomorphic with of Molluginaceae, some of which also have low seed numbers. Meanwhile, the phylogenetic data suggest that the large flowers of *Calyptrotheca* are derived within this assemblage, as small flowers characterize all possible outgroups (e.g. Smith et al., 2018; cf. Arakaki et al., 2011).

***The ACPT clade*.**[102] Current data seem to support the relation of Talinaceae as sister to a clade comprising Anacampserotaceae ("A"), Cactaceae ("C"), and Portulacaceae s. str. ("P"). The only exception is Brockington et al. (2009), which showed 98% BP for a clade including the single samples of *Talinum* and *Portulaca* as sister, with no *Anacampseros* sample.

But the relations among the families besides Talinaceae cannot be considered resolved. Nyffeler (2007) found 78% BP and 0.72 PP for the relation (P,(A,C)). This relationship also emerged in Ocampo and Mair-Sánchez (2018; cpDNA; < 0.95 PP), the problematic Smith et al. (2018; both cpDNA and nuclear loci; 55% BP), and Yao et al. (2019; cpDNA genome; 100% BP). Lendel et al. (2013) also found this relationship using only *matK*, but no support level was indicated.

Meanwhile, (A,(P,C) emerged from the partial chloroplast genome plus nuclear *phyC* (Arakaki et al., 2011; 100% BP), with several cpDNA loci plus nrDNA (Anton et al., 2014; 92% BP, 0.93 PP), and with much less support with four cpDNA loci, Ocampo and Columbus. 2010; 55% BP, 0.59 PP).

The third possible relationship, ((A,P,),C), first emerged in Hershkovitz and Zimmer (1997; nrDNA; 55% BP) and was corroborated later by Applequist and Wallace (2000; cpDNA *ndhF*; 64% BP) and Ogburn and Edwards (2015). The nuclear genome (Moore et al., 2018) and whole transcriptome analysis (Wang et al., 2018) strongly supported this relationship. In the former case, BP support was (98-)100% in all analyses. However, the loci manifested considerable incongruence, about half supporting ((A,P),C) and a quarter each supporting (A,(P,C)) and (P,(A,C)). In the transcriptome analysis, the proportions of loci supporting the latter two relations were less but not insignificant.

---

[101] I have not reviewed general evidence for ecologically-related sequence convergence of plastome genes, but I assume that this does not involve the entire plastome in a coordinated way. I also have not examined the actual data to evaluate other sorts of biased or error.

[102] There is no formal taxonomic rank between suborder and family. The ACPT clade would be legitimately named in phylogenetic nomenclature. I agree generally with Williams and Ebach's (2020) criticism of the latter concept, but clearly current traditional nomenclature has its limitations.

As in the case for Basellaceae and Didiereaceae, it appears that ACP loci phylogenies vary with respect to any fixed organismal cladogram. However, in this case, it appears that, while analyses using the entire cpDNA genome support (P,(A,C)), this relationship is counterevidenced by analyses using about half of the cpDNA genome plus *phyC* [i.e., (P,(A,C)); 100% BP] and, less so, in analysis of 1–5 cpDNA loci, sometimes in combination with a single nuclear locus. These analyses found (P,(A,C)) or (A,(P,C)), though with weaker support.

It would be useful to reanalyze especially the cpDNA loci to localize any incongruencies among these and with the genomic and transcriptomic data. Wang et al. (2018) did report "outlying"[103] loci among nuclear transcripts, one each supporting (P,(A,C)) or (A,(P,C)), but they did not find anything about these loci that suggested, e.g., errant homology. Also, again, Wang et al. (2018) provided methodological results supposedly showing evidence for ancient hybridization involving Anacampserotaceae, Cactaceae, and Portulacaceae, but they did not explain the method or results.

As discussed earlier, I find both HGT and ILS explanations to be troubling in the case of "deep" phylogenetic branches. By "deep," I mean on the order of 50+ million years, which is believed by many to be age range of the oldest Portulacineae divergences (e.g. Lendel et al., 2013; Goolsby et al., 2018a; Yao et al., 2019). Over time, mutation superimposes over the discordant phylogenetic signal, yielding an irresolvable polytomy among the descendents of the ancestral "source" and "recipient" lineages. This question will be revisited in a subsequent work on Portulacineae diversification age and interpretive consequences.

There is a paradox here. HGT and ILS events involve closely related taxa. Slowly evolving loci ought to be nearly identical in such taxa, hence HGT and ILS would not be detected. Hybridization/sorting are detected in *rapidly* evolving loci. Accordingly, it seems unlikely that evidence for HGT/ILS should persist at such loci for more than 50 million years. By this time, in rapidly evolving loci, this evidence would be overwritten by subsequent evolution. This suggests that incongruency by itself is not sufficient to diagnose HGT/ILS. Perhaps VGT provides a better explanation.

And again, the nonmolecular evidence is peculiar. Anacampserotaceae and Portulacaceae s. str. share the chromosome base number x = 9, and this differs from apparent base numbers of Talinaceae (x = 12) and Cactaceae (x = 11), as well as the outgroups of these, Basellaceae and Didiereaceae (x = 12), and Montiaceae (x = 12 in *Phemeranthus* and x = 11 among basal Cistantheae). If Anacampserotaceae and Portulacaceae s. str. are sister, x = 9 is a synapomorphy. The other configurations imply that x = 11 among Cactaceae was derived from x = 9 rather than x = 12.

These families also are *mostly* chamaephytic, herbaceous perennials with succulent leaves. These observations would seem to be in agreement with a sister relation. But detailed examination suggests that the morphological similarities among most species may represent convergence. Per Arakaki et al. (2011; cf. Smith et al., 2018), the successive outgroups of the essentially herbaceous genus *Anacampseros* are the somewhat woodier monotypic genera *Grahamia* and *Talinopsis* (100% BP each). The other molecular analyses that sampled both of these genera are Hershkovitz and Zimmer (1997), Nyffeler (2007), Ocampo and Columbus (2010), and Smith et al. (2018). These analyses indicated that these genera are not sister and that their genetic divergence is considerable. The results of these analyses support or at least are compatible with the notion of earliest divergence of the

[103] "Outliers" in phylogenomic jargon are loci whose cladograms include edges not present in the optimal MSC consensus but whose statistical support strongly biases concatenated analysis in favor of the outlier tree. Walker et al. (2018) found that with outliers removed, concatenated and MSC analyses yielded the same result.

two woody genera. They do not appear to be sister taxa, which implies that the woody condition is ancestral.

Talinopsis is a suffrutescent chamaephyte, whereas *Grahamia* is a phanerophyte, a scrambling shrub.[104] Outgroup comparison (with Talinaceae, Didiereaceae, Basellaceae, and Halophytaceae) would suggest that a more or less woody chamaephytic to phanerophytic habit is ancestral among Anacampserotaceae. In contrast to Anacampserotaceae, the ancestral form of Portulacaceae would seem to be a rather smallish herb. The most diminutive forms, e.g., *P. bicolor* and *P. quadrifida*, occur among the Australian and African "OL clade," which is sister to the larger "AL clade" (Ocampo and Columbus, 2012). The suffrutescent forms, such as *P. suffrutescens* and *P. molokiniensis*, are derived within the latter clade. Thus, regardless of how ambiguities are resolved, Portulacineae phylogeny manifests multiple transitions between herbaceous and arborescent forms.

Cactaceae phylogeny. Figure 1 defers to the cladistic interpretation Guerrero et al. (2019), which appears to be more or less a consensus based on targeted gene and phylogenomic evidence, including references cited here. This interpretation is adequate for the purposes of the present work, because proposed alternative phylogenies do not affect the present conclusions regarding the Portulacineae backbone. But I emphasize that, despite the simplicity of the reconstruction, the supporting evidence is no more robust than that for the comparatively very few taxa discussed above. This is especially evident from Wang et al. (2018: 5, Fig. 2), which shows notably small fractions of transcriptomic loci supporting clades in the Cactaceae portion of their tree, as well as Wang et al. (2018: Fig. S2), which implicates (but does not prove) a history of hybridization in the origin of major cactus clades (see also Copetti et al., 2017).

3. Summary

At the organismal level, monophyly of Portulacineae and several of taxa included therein appears to be a well-supported. By organismal level, I mean that the included taxa indeed descended from individuals that were not also organismal ancestors of other extant taxa. Nonetheless, organismal phylogenetic relations among many of the taxa have not been established convincingly. This is despite the expectations of numerous targeted gene and phylogenomic analysis. But, as I have argued elsewhere (Hershkovitz, 2019a), the expectations of cladistic clarity were themselves, at best, misguided. What I add here to this criticism is the notion that the very attempt to resolve the irresolvable reflects a constitutive cognitive tendency.

The reconciliation presented here is cursory. The already massive available genetic data for Portulacineae merit more thorough subjective (as opposed to purely computational) analysis. However, I suggest that the cladistic resolution described here will change little. This is to say that *any* new study will yield new results that might agree with those of some previous studies, but, axiomatically, not all. Do new studies serve any useful purpose? By themselves, no. They are useful only in a framework of reconciliation that explains *why* the new results agree or disagree.

In the meantime, it is more worthwhile to concentrate on the evolutionary *implications* of the currently available results. I believe that the data are consistent with PEI and not with any of the

[104] Because of its shrubby, scrambling habit and its whorl of scarious bracts subtending its flowers, Hershkovitz (1993) suggested that *Grahamia* was a "missing link" between classical Portulacaceae and certain scrambling Andean species of *Pereskia* s. lato. Molecular data disproved this hypothesis, demonstrating that Anacampserotaceae are monophyletic and that Caribbean taxa, viz. *Leuenbergeria*, are the basalmost of Cactaceae.

conventional (or even less conventional) notions of evolution. PEI, in turn, has significant implications towards general evolutionary interpretation, in particular with respect to the statistical paradigm of "phylogenetic comparative analysis" (e.g., Hershkovitz, 2018b, 2019a, b).

To the above, the present work adds an additional dimension: cognition. This is the idea that some of the apparent disagreements concerning available data might owe to factors beyond simple molecular evolutionary mechanics (e.g., HGT, ILS, VGT), methods (e.g., ML, MP, 3TS), and even beyond epistemology (e.g. evolutionary determinism versus indeterminism). Some disagreement may owe to nothing but constitutive cognitive processes that tend to canalize perceptions in certain predictable, yet not identical ways. As in the Necker Cube. Adding this dimension to systematics helps consummate the objective of reconciliation.

ACKNOWLEDGMENTS

Elaboration of this and all of my works posted since 2018 would not have been possible except for the generous support and dedicated efforts of Sandra Steuermann and Gabriela Feldman of Fundación Reshet in Chile (www.reshet.cl). Jorge Mpodozis recommended to me the Necker Cube, among various other possibilities, as an ideal demonstration of cognitive determination.

LITERATURE CITED

Abadi, S., D. Azouri, T. Pupko, and I. Mayrose. 2019. Model selection may not be a mandatory step for phylogeny reconstruction. Nature Communications 10: 934. https://doi.org/10.1038/s41467-019-08822-w

Alfaro, M.E. and M.T. Holder. 2006. The posterior and the prior in Bayesian phylogenetics. Ann. Rev. Ecol. Evol. Syst. 37: 19–42. https://doi.org/10.1146/annurev.ecolsys.37.091305.110021

Anton, A.M., T. Hernández-Hernández, J.A. De-Nova, and V. Sosa. .2014. Evaluating the phylogenetic position of the monotypic family Halophytaceae (Portulacinae [sic], Caryophyllales) based on plastid and nuclear molecular data sets. Botanical Sciences 92: 351–361. http://www.scielo.org.mx/scielo.php?script=sci_arttext&pid=S2007-42982014000300004

Applequist, W.L., W.L. Wagner, E.A. Zimmer, and M. Nepokroeff. 2006. Molecular evidence resolving the systematic position of *Hectorella* (Portulacaceae). Syst. Bot. 31: 310–319. https://doi.org/10.1600/036364406777585900

Arakaki, M., P.A. Christin, R. Nyffeler, A. Lendel, U. Eggli, R.M. Ogburn, E. Spriggs, M.J. Moore, and E.J. Edwards. 2011. Contemporaneous and recent radiations of the world's major succulent plant lineages. Proc. Nat. Acad. Sci. (USA) 108: 8379–8383. https://doi.org/10.1073/pnas.1100628108

Autzen, B. 2018. Bayesian convergence and the fair-balance paradox. Erkenntnis 83: 23–263. https://doi.org/10.1007/s10670-017-9888-0

Behnke, H.-D and T.J. Mabry (eds.). 1994. Caryophyllales, Evolution and Systematics. Springer, Berlin. https://doi.org/10.1007/978-3-642-78220-6

Brockington, S.F., R. Alexandre, J. Ramdial, M.J. Moore, S. Crawley, A. Dhingra, K. Hilu, D.E. Soltis, and P.S. Soltis. 2009. Phylogeny of the Caryophyllales sensu lato: revisiting hypotheses on pollination biology and perianth differentiation in the core Caryophyllales. Int. J. Plant Sci. 170: 627–643. https://doi.org/10.1086/597785

Cai, L., Z. Xi, A.M. Amorim, M. Sugumaran, J.S. Rest, L. Liu, and C.C. Davis. 2019. Widespread ancient whole-genome duplications in Malpighiales coincide with Eocene global climatic upheaval. New Phytol. 221: 565–576. https://doi.org/10.1111/nph.15357

Candolle, A.P. 1828. Mémoire sur la Famille des Crassulacées. Collection de Mémoires pour Servir a l'Histoire du Règne Végétal, vol. 2. Treuttel et Würtz, Paris. https://bibdigital.rjb.csic.es/records/item/12311-collection-de-memoires-second-memoire

Carlquist, S. 1998. Wood anatomy of Portulacaceae and Hectorellaceae: ecological, habital, and systematic implications. Aliso 16: 137–153. https://doi.org/10.5642/aliso.19971602.09

Carolin, R.C. 1987. A review of the family Portulacaceae. Austral. J. Bot. 35: 383–412. https://doi.org/10.1071/BT9870383

Carolin, R.C. 1993. Portulacaceae. In: K. Kubitzki, V. Bittrich, and J. Rohwer (eds.), The Families and Genera of Vascular Plants, vol. 2. Springer, Berlin, Heidelberg, New York. https://doi.org/10.1007/978-3-662-02899-5_64

Chen, D. G.J. Burleigh, and D. Fernández-Baca. 2007. Spectral partitioning of phylogenetic data sets based on compatibility. Syst. Biol. 56: 623–632. https://doi.org/10.1080/10635150701499571

Chen, M.-Y., D. Liang, and P. Zhang. 2015. Selecting question-specific genes to reduce incongruence in phylogenomics: a case study of jawed vertebrate backbone phylogeny. Syst. Biol. 64: 1104–1120. https://doi.org/10.1093/sysbio/syv059

Chippindale, P.T. and J.J. Wiens. 1994. Weighting, partitioning, and combining characters in phylogenetic analysis. Syst. Biol. 43: 278–287. https://doi.org/10.1093/sysbio/43.2.278

Copetti, D., A. Búrquez, E. Bustamante, J.L.M. Charboneau, K.L. Childs, L.E. Eguiarte, S. Lee, T.L. Liu, M.M. McMahon, N.K. Whiteman, R.A. Wing, M.F. Wojciechowski, and M.J. Sanderson. 2017. Extensive gene tree discordance and hemiplasy shaped the genomes of North American columnar cacti. Proc. Nat. Acad. Sci. USA 114: 12003–12008. https://doi.org/10.1073/pnas.1706367114

Cronquist, A. and A.L. Takhtadzhian. 1981. An integrated system of classification of flowering plants. Columbia Univ. Press, New York, New York..

Darwin, C. 1859. On the Origin of Species by Means of Natural Selection, or the Preservation of Favoured Races in the Struggle for Life. John Murray, London, UK). In: J. van Wyhe (ed.). The Complete Work of Charles Darwin Online. http://darwin.online.org.uk

Denk T., G.W. Grimm, P.S. Manos, M. Deng, and A.L. Hipp. 2017 An updated infrageneric classification of the oaks: review of previous taxonomic schemes and synthesis of evolutionary patterns. In: E. Gil-Pelegrín, J. Peguero-Pina, and D. Sancho-Knapik D. (eds.), Oaks Physiological Ecology. Exploring the Functional Diversity of Genus *Quercus* L.. Tree Physiology, vol. 7. Springer, Cham. https://doi.org/10.1007/978-3-319-69099-5_2

Edwards, E.J. 2019. Evolutionary trajectories, accessibility and other metaphors: the case of C4 and CAM photosynthesis. New Phytol. 223: 1742–1755. https://doi.org/10.1111/nph.15851

Edwards, E.J. and R.M. Ogburn. 2012. Angiosperm responses to a low-CO2 world: CAM and C4 photosynthesis as parallel evolutionary trajectories. Int. J. Plant Sci. 173: 724–733. https://doi.org/10.1086/666098

Escobar, J. S., S. Glémin, and N. Galtier. 2011. GC-biased gene conversion impacts ribosomal DNA evolution in vertebrates, angiosperms, and other eukaryotes. Molecular Biology and Evolution, 28(9), 2561–2575. https://doi.org/10.1093/molbev/msr079

García-Sandoval, R. 2014. Why some clades have low bootstrap frequencies and high Bayesian posterior probabilities. J. Israel Ecol. Evol. 60: 41–44. https://doi.org/10.1080/15659801.2014.937900

Gilman, I.S., J.J. Moreno-Villena, Z.R. Lewis, E.W. Goolsby, and E.J. Edwards. 2021. Gene co-expression reveals the modularity and integration of C4 and CAM in Portulaca. bioRxiv. https://doi.org/10.1101/2021.07.07.451465

Goloboff, P.A., A. Torres Galvis, and J.S. Arias. 2018. Parsimony and model-based phylogenetic methods for morphological data: comments on O'Reilly et al. Palaeontology 61: 625–630. https://doi.org/10.1111/pala.12353

Goolsby, E.W., A.J. Moore, L.P. Hancock, J.M. de Vos, and E.J. Edwards. 2018a. Molecular evolution of key metabolic genes during transitions to C4 and CAM photosynthesis. Amer. J. Bot. 105: 602–613. https://doi.org/10.1002/ajb2.1051

Goolsby, E.W., A.J. Moore, L.P. Hancock, J.M. de Vos, and E.J. Edwards. 2018b. Data from: Molecular evolution of key metabolic genes during transitions to C4 and CAM photosynthesis. Dryad Digital Repository. https://doi.org/10.5061/dryad.47m18

Grünwald, P. and T. van Ommen. 2017. Inconsistency of Bayesian inference for misspecified linear models, and a proposal for repairing it. Bayesian Analysis 12: 1069–1103. https://doi.org/10.1214/17-BA1085

Guerrero P.C., L.C. Majure, A. Cornejo-Romero, and T. Hernández-Hernández T. 2019. Phylogenetic relationships and evolutionary trends in the cactus family. J. Heredity 110: 4–21. https://doi.org/10.3372/wi.51.51208

Harpke, D. and A. Peterson. 2006. Non-concerted ITS evolution in *Mammillaria* (Cactaceae). Mol. Phylogen. Evol. 41: 579–593. https://doi.org/10.1016/j.ympev.2006.05.036

Harpke, D. and A. Peterson. 2007. Quantitative PCR revealed a minority of ITS copies to be functional in *Mammillaria* (Cactaceae). Int. J. Plant Sci. 168: 1157–1160. https://doi.org/10.1086/520729

Harpke, D. and A. Peterson. 2008. Extensive 5.8 S nrDNA polymorphism in *Mammillaria* (Cactaceae) with special reference to the identification of pseudogenic internal transcribed spacer regions. J. Plant Res. 121: 261–270. https://doi.org/10.1007/s10265-008-0156-x

Hipp, A.L., P.S. Manos, M. Hahn, M. Avishai, C. Bodénès, J. Cavender-Bares, A.A. Crowl, M. Deng, T. Denk, S. Fitz-Gibbon, O. Gailing, M.S. González-Elizondo, A. González-Rodríguez, G.W. Grimm, X.-L. Jiang, A. Kremer, I. Lesur, J.D. McVay, C. Plomion, H. Rodríguez-Correa, E.-D. Schulze, M.C. Simeone, V.L. Sork, and S. Valencia-Avalos. 2020. Genomic landscape of the global oak phylogeny. New Phytol. 226: 1198–1212. https://doi.org/10.1111/nph.16162

Hernández-Ledesma, P., W.G. Berendsohn, T. Borsch, S. von Mering, H. Akhani, S. Arias, I. Castañeda-Noa, U. Eggli, R. Eriksson, H. Flores-Olvera, S. Fuentes-Bazán, G. Kadereit, C. Klak, N. Korotkova, R. Nyffeler, G. Ocampo, H. Ochoterena, B. Oxelman, R.K. Rabeler, A. Sanchez, B.O. Schlumpberger, and P. Uotila. 2015. A taxonomic backbone for the global synthesis of species diversity in the angiosperm order Caryophyllales. Willdenowia 45: 281–383. https://doi.org/10.3372/wi.45.45301

Hershkovitz, M.A. 1993. Revised circumscription and sectional taxonomy of *Calandrinia* Kunth and *Montiopsis* Kuntze (Portulacaceae) with notes on phylogeny of the portulacaceous alliance. Ann. Missouri Bot. Gard. 80: 333–365. https://doi.org/10.2307/2399789

Hershkovitz, M.A. 2006. Evolution of western American Portulacaceae in the Andean region. Gayana 63: 13–74. http://dx.doi.org/10.4067/S0717-66432006000100002

Hershkovitz, M.A. 2018a. Perspectives from Montiaceae (Portulacineae) evolution. I. Phylogeny and phylogeography. Preprints 2018 2018090096. https://doi.org/10.20944/preprints201809.0096.v2

Hershkovitz, M.A. 2018b. Perspectives from Montiaceae (Portulacineae) evolution. II. Ecological evolution, phylogenetic comparative analysis, and the principle of evolutionary idiosyncraticity. Preprints 2018 2018090566. https://doi.org/10.20944/preprints201809.0566.v2

Hershkovitz, M.A. 2019a. Systematics, Evolution, and Phylogeography of Montiaceae (Portulacineae). Phytoneuron 2019-27: 1-77. http://www.phytoneuron.net/2019Phytoneuron/27PhytoN-Montiaceae.pdf

Hershkovitz, MA. 2019b. The 'Holy Grail' in phylogenetic reconstruction: seeing the forest for the trees? EcoEvoRxiv. https://doi.org/10.32942/osf.io/b9mtn

Hershkovitz, M.A. 2020a. Systematics of *Calandrinia pilosiuscula* DC a.k.a. *Calandrinia compressa* Schrad. ex DC (Montiaceae–Montioideae). EcoEvoRxiv. https://doi.org/10.32942/osf.io/wgaf3

Hershkovitz, M.A. 2020b. *Calandrinia jompomae* (MONTIACEAE), another overlooked species in the Chilean flora. EcoEvoRxiv. https://doi.org/10.32942/osf.io/gv5wr

Hershkovitz, M.A. 2020c. Bertero's ghost revisited: new typifications of *Talinum linaria* Colla and *Calandrinia gaudichaudii* Barnéoud (= *Calandrinia pilosiuscula* DC; Montiaceae). EcoEvoRxiv. https://doi.org/10.32942/osf.io/n4d5j

Hershkovitz, M.A. 2020d. *Rumicastrum* Ulbrich (Montiaceae): a beautiful name for the Australian calandrinias. Phytologia 102: 116–123. https://www.phytologia.org/uploads/2/3/4/2/23422706/102_3_116-123hershkovitzrumicasterum9-9-20final.pdf

Hershkovitz, M.A. 2021a. Corrections to Phytologia,vol. 102 (3) – *Rumicastrum dielsii* (Poelln.) Carolin is an "alien," *R. cylindricum* (Poelln.) Carolin is Elatinaceae, *R. monogynum* (Poelln.) Carolin is incertae sedis, and the combination in *Rumicastrum* for *Talinum nanum* Nees was neglected. EcoEvoRxiv. https://doi.org/10.32942/osf.io/g6s3a

Hershkovitz, M.A. 2021b. Evolutionary insights from DNA sequences from Chaetanthera Ruiz & Pav. and Oriastrum Poepp. & Endl. (Asteraceae; Mutisieae). I. Of molecules and systematics. EcoEvoRxiv. https://doi.org/10.32942/osf.io/ak68m/

Hershkovitz, M.A., E.A. Zimmer, and W.J. Hahn. 1999. Ribosomal DNA and angiosperm evolution. In: P.M. Hollingsworth, R.M. Bateman, and R.J. Gornall (eds.), Molecular Systematics and Plant Evolution. Taylor & Francis, London.

Hershkovitz, M.A. and E.A. Zimmer. 1997. On the evolutionary origins of the cacti. Taxon 46: 217–242. https://doi.org/10.2307/1224092

Hershkovitz, M.A. and E.A. Zimmer. 2000. Ribosomal DNA evidence and disjunctions of western American Portulacaceae. Mol. Phylogen. Evol. 15: 419–439. https://doi.org/10.1006/mpev.1999.0720

Holtum, J.A., L.P. Hancock, E.J. Edwards, and K. Winter. 2017. Optional use of CAM photosynthesis in two C4 species, *Portulaca cyclophylla* and *Portulaca digyna*. J. Plant Physiol. 214: 91–96.

Holtum, J.A., L.P. Hancock, E.J. Edwards, and K. Winter. 2018. Crassulacean acid metabolism in the Basellaceae (Caryophyllales). Plant Biology 20: 409–414. https://doi.org/10.1111/plb.12698

Huelsenbeck, J.P. C. Ané, B. Larget, and F. Ronquist. 2008. A Bayesian perspective on a non-parsimonious parsimony model. Syst. Biol. 57: 406–419, https://doi.org/10.1080/10635150802166046

Inkpen, S.A. and W.F. Doolittle. 2016. Molecular phylogenetics and the perennial problem of homology. J. Mol. Evol. 83: 184–192. https://doi.org/10.1007/s00239-016-9766-4

Kremer, A. and A.L. Hipp. 2020. Oaks: an evolutionary success story. New Phytol. 226: 987–1011. https://doi.org/10.1111/nph.16274.

Lendel, A. 2013. South American Cacti in time and space: studies on the diversification of the tribe Cereeae, with particular focus on subtribe Trichocereinae (Cactaceae). PhD Dissertation, Univ. of Zurich, Zurich, Switzerland. https://www.zora.uzh.ch/id/eprint/93287/1/20141906.pdf

Lendel, A., A. Antonelli, and R. Nyffeler. 2013. Tackling the molecular dating paradox: underestimated pitfalls and best strategies when fossils are scarce. In: A. Lendel, South American Cacti in time and space: studies on the diversification of the tribe Cereeae, with particular focus on subtribe Trichocereinae (Cactaceae). PhD Dissertation, Univ. of Zurich, Zurich, Switzerland. https://www.zora.uzh.ch/id/eprint/93287/1/20141906.pdf

Leonelli, S. 2013. Classificatory theory in biology. Biol. Theory 7: 338–345. https://doi.org/10.1007/s13752-012-0049-z

Lewis, P.O., M.-H. Chen, L. Kuo, L.A. Lewis, K. Fučíková, S. Neupane, Y.-B. Wang, and D. Shi. 2016. Estimating Bayesian phylogenetic information content. Syst. Biol. 65: 1009–1023, https://doi.org/10.1093/sysbio/syw042

Linkem, C.W., V.N. Minin, and A.D. Leaché. 2016. Detecting the anomaly zone in species trees and evidence for a misleading signal in higher-level skink phylogeny (Squamata: Scincidae)., Syst. Biol. 65: 465–477. https://doi.org/10.1093/sysbio/syw001

Lunerová, J., S. Renny-Byfield, R. Matyášek, A. Leitch, and A. Kovařík. 2017. Concerted evolution rapidly eliminates sequence variation in rDNA coding regions but not in intergenic spacers in *Nicotiana tabacum* allotetraploid. Plant Syst. Evol. 303: 1043–1060. https://doi.org/10.1007/s00606-017-1442-7

Maturana, H.R. and J. Mpodozis. 2000. The origin of species by means of natural drift. Rev. Chil. Hist. Nat. 73: 261–310. https://www.researchgate.net/profile/Jorge_Mpodozis/publication/262497422_El_origen_de_las_especies_por_medio_de_la_deriva_natural/links/0c96053bd4f4696eb5000000.pdf

Maturana, H. and F.J. Varela. 1992. The Tree of Knowledge: The Biological Roots of Human Understanding. *(*revised ed., translated by R. Paolucci). Shambhala Publications, Boulder, CO.

McNeill, J. 1974. Synopsis of a revised classification of the Portulacaceae. Taxon 23: 725–728. https://doi.org/10.2307/1218433

Moore, A.J., J.M. de Vos, L.P. Hancock, E. Goolsby, and E.J. Edwards. 2017. Data from: Targeted enrichment of large gene families for phylogenetic inference: phylogeny and molecular evolution of photosynthesis genes in the Portullugo clade (Caryophyllales), Dryad, Dataset, https://doi.org/10.5061/dryad.7h3f6

Moore, A.J., J.M. de Vos, L.P. Hancock, E. Goolsby, and E.J. Edwards. 2018. Targeted enrichment of large gene families for phylogenetic inference: phylogeny and molecular evolution of photosynthesis genes in the Portullugo clade (Caryophyllales). Syst. Biol. 67: 367–383. https://doi.org/10.1093/sysbio/syx078

Nelson, G. and N. Platnick. 1981. Systematics and Biogeography. Columbia Univ. Press, New York, New York. https://www.researchgate.net/publication/258384023_Systematics_and_Biogeography_Cladistics_and_Vicariance

Neupane, S., K. Fučíková, L.A. Lewis, L. Kuo, M.-H. Chen, and P.O. Lewis. 2019. Assessing combinability of phylogenomic data using Bayes factors. Syst. Biol. 68: 744–754, https://doi.org/10.1093/sysbio/syz007

Nyffeler, R. 2007. The closest relatives of cacti: insights from phylogenetic analyses of chloroplast and mitochondrial sequences with special emphasis on relationships in the tribe Anacampseroteae. Amer. J. Bot., 94: 89–101. https://doi.org/10.3732/ajb.94.1.89

Nyffeler, R. and U. Eggli. 2010. Disintegrating Portulacaceae: a new familial classification of the suborder Portulacineae (Caryophyllales) based on molecular and morphological data. Taxon 59: 227–240. https://doi.org/10.10027tax.591021

Ocampo, G. and J.T. Columbus. 2010. Molecular phylogenetics of suborder Cactineae (Caryophyllales), including insights into photosynthetic diversification and historical biogeography. Amer. J. Bot. 97: 1827–1847. https://doi.org/10.3732/ajb.1000227

Ocampo, G. and J.T. Columbus. 2012. Molecular phylogenetics, historical biogeography, and chromosome evolution of *Portulaca* (Portulacaceae). Mol. Phylogen. Evol. 63: 97–112. https://doi.org/10.1016/j.ympev.2011.12.017

Ogburn, R.M. and E.J. Edwards. 2015. Life history lability underlies rapid climate niche evolution in the angiosperm clade Montiaceae. Mol. Phylogen. Evol. 92: 181–192. https://doi.org/10.1016/j.ympev.2015.06.006

Richards, R.J. 2012. Darwin's principles of divergence and natural selection: why Fodor was almost right. Stud. Hist. Philos. Biol. Biomed. Sci. 43: 256–68. https://doi.org/10.1016/j.shpsc.2011.10.014

Rosenberg, N.A. 2013. Discordance of species trees with their most likely gene trees: a unifying principle. Mol. Biol. Evol. 30: 2709–2713. https://doi.org/10.1093/molbev/mst160

Smith, S.A., J.W. Brown, Y. Yang, R. Bruenn, C.P. Drummond, S.F. Brockington, J.F. Walker, N. Last, N.A. Douglas, and M.J. Moore. 2017. Data from: Disparity, diversity, and duplications in the Caryophyllales, Dryad, Dataset, https://doi.org/10.5061/dryad.jf7np

Smith, S.A., J.W. Brown, Y. Yang, R. Bruenn, C.P. Drummond, S.F. Brockington, J.F. Walker, N. Last, N.A. Douglas, and M.J. Moore. 2018. Disparity, diversity, and duplications in the Caryophyllales. New Phytol. 217: 836–854. https://doi.org/10.1111/nph.14772

Soltis, D.[E.] and P.[S.] Soltis. 2018. The Great Tree of Life. Academic Press, London, UK. https://doi.org/10.1016/C2016-0-03440-X

Ronse de Craene, L.P. 2013. Reevaluation of the perianth and androecium in Caryophyllales: implications for flower evolution. Plant Syst. Evol. 299: 1599–1636. https://doi.org/10.1007/s00606-013-0910-y

Sullivan, A.R., B. Schiffthaler, S.L. Thompson, N.R. Street, and X.-R. Wang, 2017. Interspecific plastome recombination reflects ancient reticulate evolution in *Picea* (Pinaceae). Mol. Biol. Evol. 34: 1689–1701. https://doi.org/10.1093/molbev/msx111

Swofford, D.L., G.J. Olsen, P.J. Waddell, and D.M. Hillis. 1996. Phylogenetic inference. In: D.M. Hillis, C. Moritz, and B.K. Mable (eds.), Molecular Systematics, 2nd ed. Sinauer, Sunderland, Massachusetts. [no digital version available]

Symonová, R. 2019. Integrative rDNAomics – importance of the oldest repetitive fraction of the eukaryote genome. Genes 10: 345. https://doi.org/10.3390/genes10050345

Thulin, M., A.J. Moore, H. El-Seedi, A. Larsson, P. Christin, P. and E.J. Edwards. 2016. Phylogeny and generic delimitation in Molluginaceae, new pigment data in Caryophyllales, and the new family Corbichoniaceae. Taxon, 65: 775–793. https://doi.org/10.12705/654.6

Varela, F. J., E. Thompson, and E. Rosch. 1992. The Embodied Mind: Cognitive Science and Human Experience. MIT Press, Cambridge, MA.

Vaschetto, L. M. and N. Ortiz. 2019. The role of sequence duplication in transcriptional regulation and genome evolution. Curr. Genomics 20: 405–408. https://doi.org/10.2174/1389202920666190320140721

Vázquez-Pardo, F., C. Pinto-Gomes, C. Vila-Viçosa, M. Ángel-Pineda, D. Garcia-Alonso, F. Márquez, and J. Blanco-Salas, J. 2015. Three new oak hybrids from southwest Iberia (Spain and Portugal). J. Intl. Oak Soc. 26: 43–56. https://dspace.uevora.pt/rdpc/bitstream/10174/18062/1/New%20Oak%20Hybrids.pdf

Virgo, N. 2019. The necessity of extended autopoiesis. Adaptive Behavior. 617. https://doi.org/10.1177/1059712319841557

Walker, J.F., Y. Yang, T. Feng, A. Timoneda, J. Mikenas, V. Hutchison, C. Edwards, N. Wang, S. Ahluwalia, J. Olivieri, N. Walker-Hale, L.C. Majure, R. Puente, G. Kadereit, M. Lauterbach, U. Eggli, H. Flores-Olvera, H. Ochoterena, S.F. Brockington, M.J. Moore, and S.A. Smith. 2018. From cacti to carnivores: improved phylotranscriptomic sampling and hierarchical homology inference provide further insight into the evolution of Caryophyllales. Amer. J. Bot. 105: 446–462. https://doi.org/10.1002/ajb2.1069

Wang, N., Y. Yang, M.J. Moore, S.F. Brockington, J.F. Walker, J.W. Brown, B. Liang, T. Feng, C. Edwards, J. Mikenas, J. Olivieri, V. Hutchison, A. Timoneda, T. Stoughton, R. Puente, L.C. Majure, U. Eggli and S.A. Smith. 2019. Evolution of Portulacineae marked by gene tree conflict and gene family expansion associated with adaptation to harsh environments. Mol. Biol. Evol. 36: 112–126, https://doi.org/10.1093/molbev/msy200

Wiens, J.J. and P.T. Chippindale. 1994. Combining and weighting characters and the prior agreement approach revisited. Syst. Biol. 43: 564–566. https://doi.org/10.1093/sysbio/43.4.564

Williams, D.M. and C. Ebach. 2020. Cladistics, a Guide to Biological Classification, ed. 3. Cambridge Univ. Press, Cambridge, UK. https://doi.org/10.1017/9781139047678

Wilson, E.O. 1998. Consilience: the Unity of Knowledge. Alfred A. Knopf, New York, New York.

Winter, K. 2019. Ecophysiology of constitutive and facultative CAM photosynthesis. J. Experimental Bot. 70: 6495–6508. https://doi.org/10.1093/jxb/erz002

Xu, B. and Z. Yang. 2016. Challenges in species tree estimation under the multispecies coalescent model. Genetics 204: 1353–1368. https://doi.org/10.1534/genetics.116.190173

Yang, Y., M.J. Moore, S.F. Brockington, J. Mikenas, J. Olivieri, J.F. Walker, and S.A. Smith. 2018. Improved transcriptome sampling pinpoints 26 ancient and more recent polyploidy events in Caryophyllales, including two allopolyploidy events. New Phytol. 217: 855–870. https://doi.org/10.1111/nph.14812

Yang, Z. and T. Zhu. 2018. Bayesian selection of misspecified models is overconfident and may cause spurious posterior probabilities for phylogenetic trees. Proc. Nat. Acad. Sci. USA 115: 1854–1859. https://doi.org/10.1073/pnas.1712673115

Yao, G., J.-J. Jin, H.-T. Li, J.-B. Yang, V.S. Mandala, M. Croley, R. Mostow, N.A. Douglas, M.W. Chase, M.J.M. Christenhusz, D.E. Soltis, P.S. Soltis, S.A. Smith, S.F. Brockington, M.J. Moore, T.-S. Yi, and D.-Z. Li. 2019. Plastid phylogenomic insights into the evolution of Caryophyllales. Mol. Phylogen. Evol. 134: 74–86. https://doi.org/10.1016/j.ympev.2018.12.023

Zhang, M., Y.-W. Tang, Y. Xu, T. Yonezawa, Y. Shao, Y.-G. Wang, Z.-P. Song, J. Yang, and W.-J. Zhang. 2019. Concerted and birth-and-death evolution of 26S ribosomal DNA in *Camellia* L. Annals Bot. 127: 63–73, https://doi.org/10.1093/aob/mcaa169

Figure captions

Fig. 1. Generalized cladogram of backbone relations among Portulacineae. This figure is taken directly from a companion work in preparation. See text for explanation of the cladistic parameter. Taxa pertinent to Anacampserotaceae, Cactaceae, and Montiaceae are indicated in aqua, lavender, and pink labels. The dots on the right side refer to developmental fate of functional sepals/petals or, in the case of Cactaceae, outermost (bracteoid to sepaloid) and innermost (petaloid) organs. Green indicates persistent and orange indicated deciduous.

Fig. 2. The Necker Cube. The top figure shows the simple planar drawing of the Necker Cube. The lower two figures emphasize the two possible cube orientations that can be perceived. (By Gauravjuvekar - Own work, CC0, https://commons.wikimedia.org/w/index.php?curid=17057834)

Fig. 3. Candolle´s notion of natural relations among Crassulaceae. From Candolle (1828: Pl. 2). The text emphasizes the ambiguity of cladistic relations suggested for genera included in the main plane of the four quadrants, which Candolle (1828) classified in "Crassulaceae Legitimae." Two "most parsimonious" cladistic resolutions are possible (i.e., it is a "Necker Cube"). It is not clear why Candolle (1828) adopted one and not the other in his classification. Williams and Ebach (2020) either overlooked or ignored this point. But the diagram manifests additional complexities also overlooked or ignored by Williams and Ebach (2020). The two genera that Candolle (1828) classified in "Crassulaceae Anomale" appear to be elevated above the plane of the upper right quadrant, and one of these is connected with another of Candolle's families, "Saxifrageae." Meanwhile, a genus of "Crassulaceae Legitimae" in the upper left quadrant is connected to another family, "Paronychieae." Designating either of these segregated families as the "outgroup" would require that the other be included in Crassulaceae. In addition, genera within the upper left and lower right quadrants are differentially displaced towards the center or periphery. This differential displacement is not reflected in Candolle's (1828) generic classification or mentioned by Williams and Ebach (2020).

Fig. 1

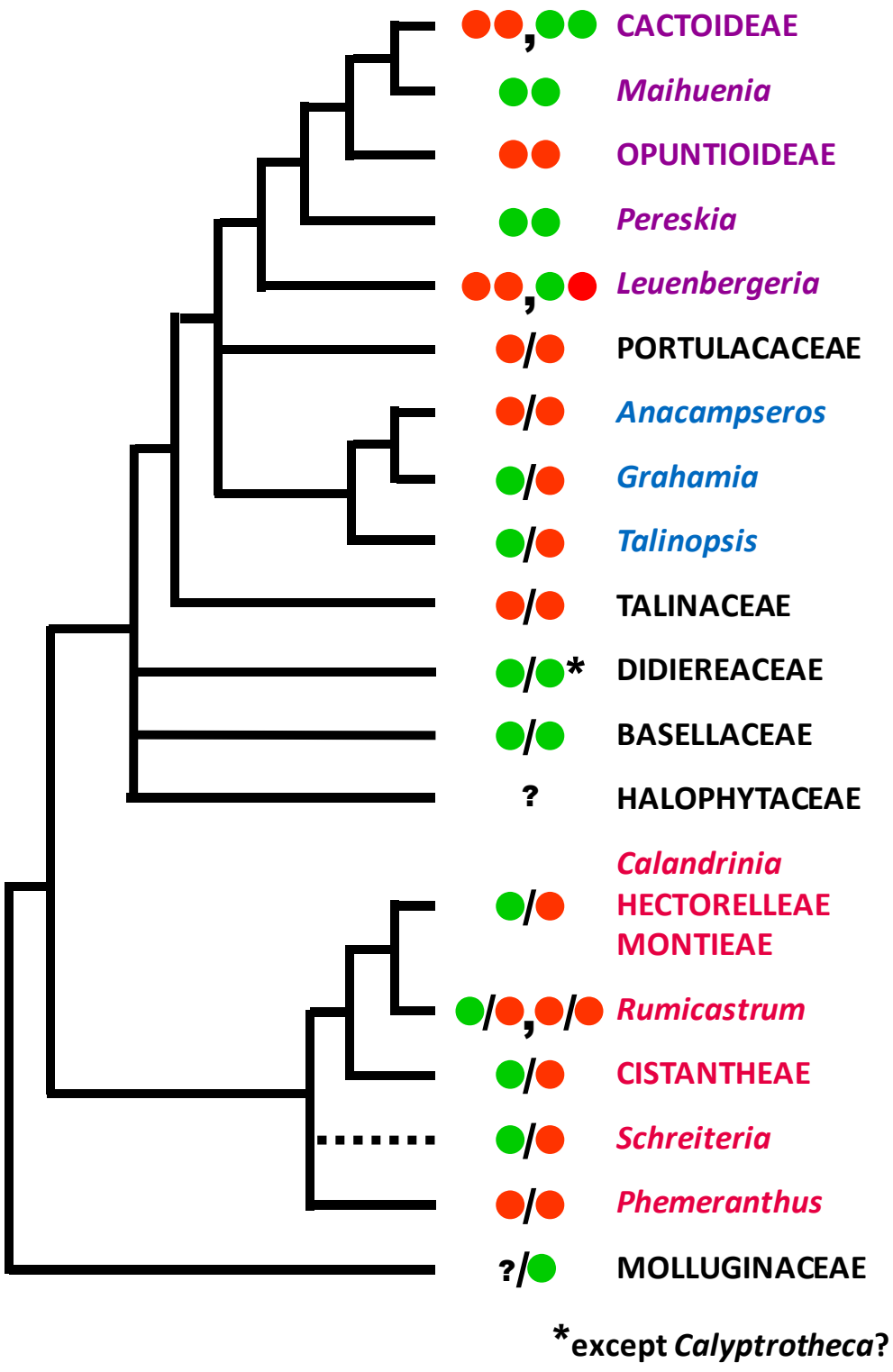

CACTOIDEAE
Maihuenia
OPUNTIOIDEAE
Pereskia
Leuenbergeria
PORTULACACEAE
Anacampseros
Grahamia
Talinopsis
TALINACEAE
DIDIEREACEAE
BASELLACEAE
HALOPHYTACEAE
Calandrinia
HECTORELLEAE
MONTIEAE
Rumicastrum
CISTANTHEAE
Schreiteria
Phemeranthus
MOLLUGINACEAE
*except Calyptrotheca?

Fig. 2

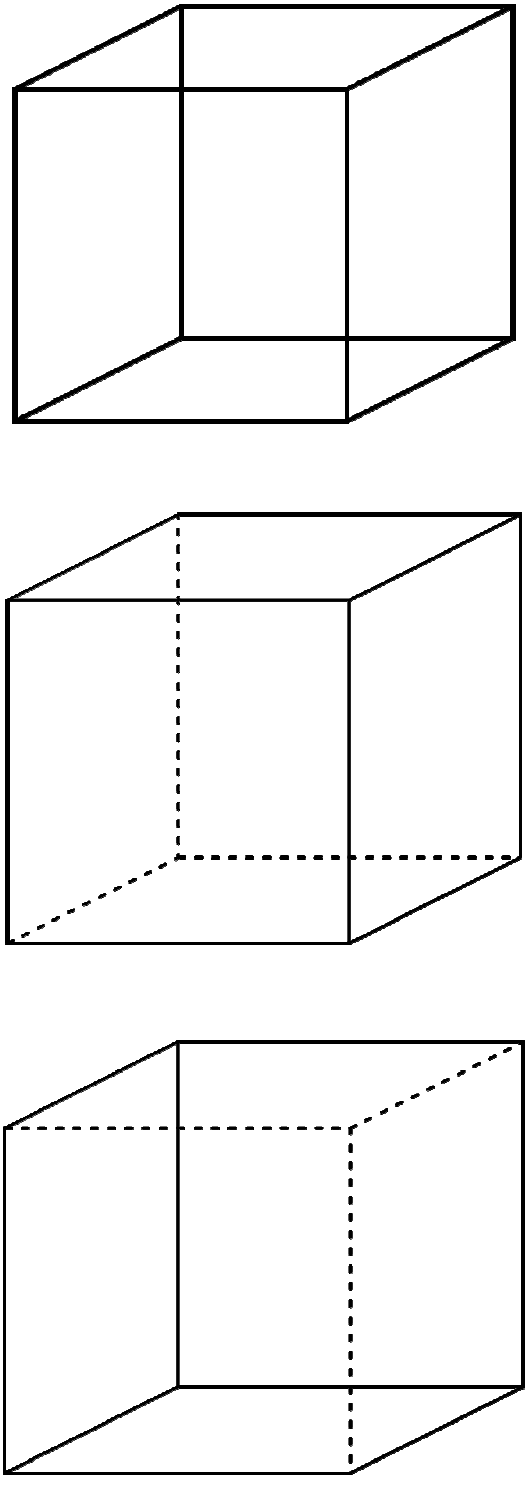
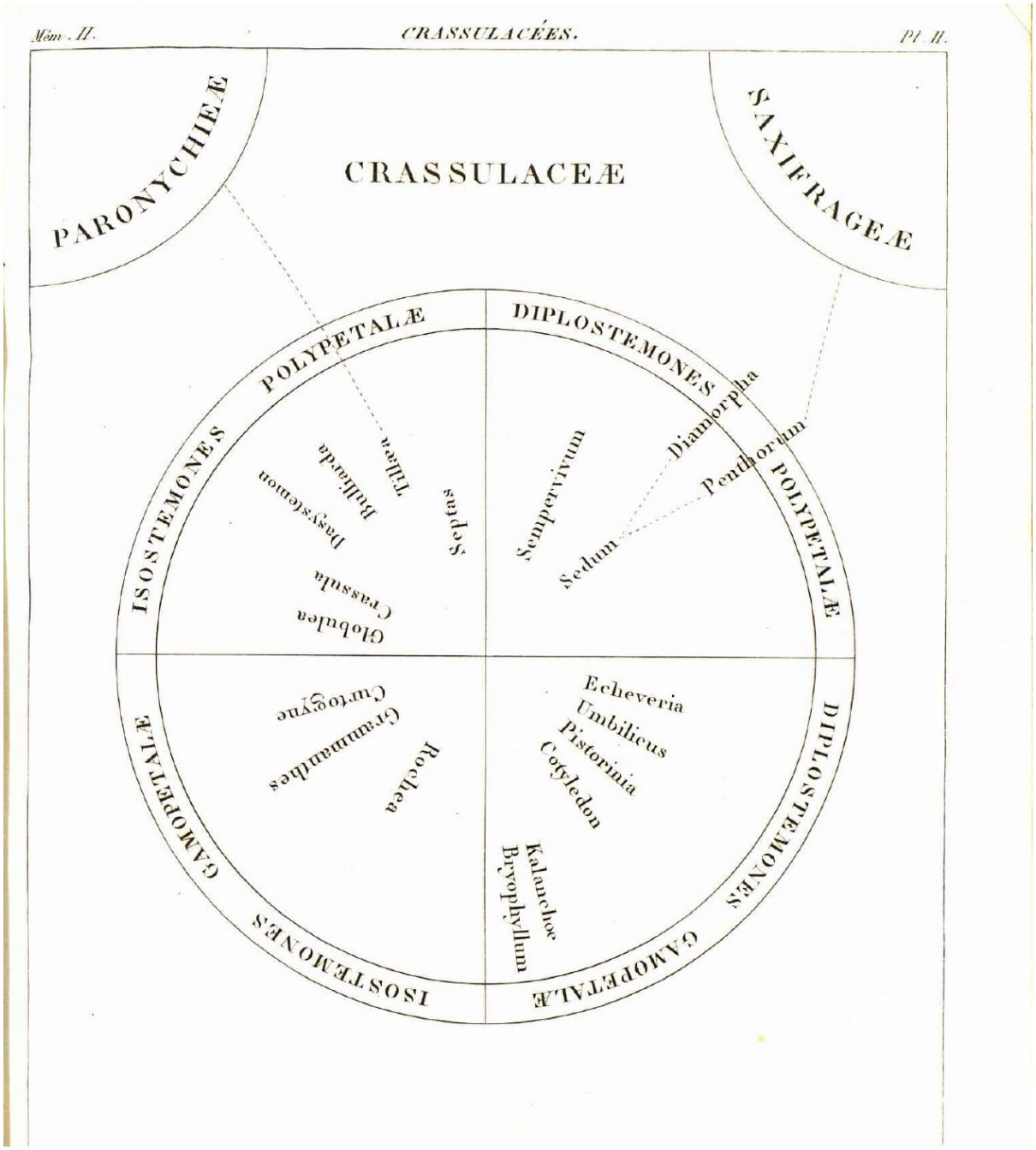


Fig. 3


Mém. II.
CRASSULACÉES.
Pl. II.
PARONYCHIEÆ
CRASSULACEÆ
SAXIFRAGEÆ
POLYPETALÆ
DIPLOSTEMONES
ISOSTEMONES
Tillæa
Bullardia
Dasystemon
Crassula
Globulea
Sedum
Sempervivum
Diamorpha
Penthorum
Sedum
Echeveria
Umbilicus
Pistorinia
Cotyledon
Kalanchoe
Bryophyllum
Rochea
Crassula
Grammanthes
Curtogyne