# Revision of a revised phylogenetic analysis of *Leucheria* Lag. sensu lato (Asteraceae; Nassauvieae) with remarks on theoretical aspects of phylogenetics and phylogenomics

Mark A. Hershkovitz
El Quisco, Chile
cistanthe@gmail.com

## ABSTRACT

Presented here is a revised analysis of a previous phylogenetic analysis of the *Leucheria* Lag. s. l. crown group (Asteraceae; Nassauvieae), which was, in turn, a reanalysis of an earlier analysis of *Leucheria* s. stricto. *Leucheria* s. l. includes *Polyachyrus* Lag, and the crown group includes *Oxyphyllum* Phil. The present analysis adds several more recently published sequences of the nuclear ribosomal DNA internal transcribed spacer region (ITS), plastome (cpDNA) *rpl32-trnL*(UAG) intergenic spacer (*rpl32-trnL*), and *trnL*(UAA) intron plus *trnL*(UAA)-*trnF*(GAA) intergenic spacer (*trnL-trnF*). While this revised analysis adds little to existing phylogenetic resolution, it provides considerable refinement and/or corroboration of existing evidence for phylogenetic relations and species taxonomy. The evidence also provides a platform for timely review of the mathematical nature of phylogenetic trees, data set incongruence, data support, operations of phylogenetic reconstructive methods, the relation between organisms and genomes, and biological and evolutionary epistemology.
**Key words:** *Leucheria*, *Polyachyrus*, *Oxyphyllum*, Asteraceae, Mutisioideae, Mutisieae, Nassauvieae, phylogenetic trees, data set incongruence, data support, epigenesis.

[CITATION: Hershkovitz, M.A. 2025. Revision of a revised phylogenetic analysis of *Leucheria* Lag. sensu lato (Asteraceae; Nassauvieae) with remarks on theoretical aspects of phylogenetics and phylogenomics. EcoEvoRxiv. https://doi.org/10.32942/X2HK9J]

## Introduction

Hershkovitz (2024a–c; see also Hershkovitz, 2024d, e) published molecular phylogenetic analyses of *Leucheria* Lag. (sensu Katinas et al., 2022 [hereafter Katinas2022]) and related Nassauvieae (Asteraceae; Mutisioideae). *Leucheria* (sensu Katinas2022) is a Patagonian genus of 28–48 mostly herbaceous species diversified in seasonally cool to cold habitats of southern South America (Jara-Arancio et al., 2017a [hereafter Jara2017]; Lavandero et al., 2020, 2024 [hereafter Lavandero2020 and Lavandero2024]; Katinas2022). These analyses were based on DNA sequences from three loci, the nuclear ribosomal DNA internal transcribed spacer region (hereafter ITS) and plastome (cpDNA) sequences of the *rpl32-trnL*(UAG)[1] intergenic spacer (hereafter *rpl32-trnL*) and the *trnL*(UAA) intron plus *trnL*(UAA)-*trnF*(GAA) intergenic spacer (hereafter *trnL-trnF*).

Hershkovitz (2024a–c) reported, variously, that:

**(i)** a previously published phylogenetic analysis of *Leucheria*/Nassauvieae by Jara2017 used the same loci but severely erroneous DNA sequences/alignments;

**(ii)** at least one species in Jara2017 was misidentified, and DNA sequences for at least two species were partially contaminant sequences;

[1] Erroneously written as "*trnL*(UAA) in Hershkovitz (2024a–c).

**(iii)** the largely lowland and mostly suffrutescent Chilean genus *Polyachyrus* Lag. is phylogenetically nested within *Leucheria* and can be recognized as *L.* sect. *Polyachyrus* (Lag.) Hershk. (Hershkovitz, 2024e);

**(iv)** the monotypic Chilean shrubby desert genus *Oxyphyllum* Phil. pertains to the monophyletic *Leucheria* crown group,[2] but its precise relation therein remained unresolved;

**(v)** the above notwithstanding, the four *Leucheria* clades recognized by Jara2017 were (more accurately and convincingly) corroborated (and later were classified as *Leucheria* sections by Hershkovitz (2024e);[3,4]

**(vi)** the phylogenetic relations among the sections named in Hershkovitz (2024e) remained unresolved, but the evidence suggests that *L.* sections *Cassiopea* and *Polyachyrus* are sister;

**(vii)** setting aside *L.* sect. *Polyachyrus* and *Oxyphyllum*, the analysis did not corroborate a sister relation between *L.* sect. *Lasiorrhiza* and remaining *Leucheria*, as reported by Jara2017, *presumed* by Pérez2020 (including Jara-Arancio and Lavandero), and later "corroborated" by Lavandero2020 (including Pérez) with 100% ML bootstrap support (BP);[5]

**(viii)** likewise, polyphyly of the genotypes of annual[6] species lumped into *L. tomentosa* (Less.) Crisci by Apodaca et al. (2021) and Katinas2022 was corroborated, but the actual number of life history transitions was not resolved;

**(ix)** *L.* sect. *Lasiorrhiza* comprises two polytypic clades that probably hybridized at some point;

**(x)** the ancestral habitat of *Leucheria* was arid/warm rather than humid/cold, contra Jara2017 and Pérez2020;

---

[2] Monophyly of this group was first reported by Luebert et al. (2009; see also Sancho et al., 2018), but that work sampled only two *Leucheria* species.

[3] Hershkovitz erroneously listed "*Leucheria* sect. *Cassiopea* D.Don, Trans. Linn. Soc. London 16(2): 215. 1830." This is essentially as Candolle (1838) wrote it. The correct name and authority is: *Leucheria* sect. *Cassiopea* (D.Don) Hook., Companion Bot. Mag. 1: 36. 1835. As noted by Katinas2022, numerous 19th Century authors erroneously referred to "*Cassiopea* D.Don" as a generic name.

[4] The clades formally classified as sections in Hershkovitz (2024e) were named informally but *inconsistently* in Jara2017, Pérez2020, and Lavandero2020, as detailed later.

[5] Put another way, without including *L.* sect. *Polyachyrus* and *Oxyphyllum* sequences, Lavandero2020 reported 100% BP support for a clade comprising all caulescent traditional *Leucheria* taxa, viz. *L.* sections *Cassiopea*, *Leucheria*, and *Macrobotrys*.

[6] Based on the presence of a vascular cambium (viz. "secondary growth"), Apodaca et al. (2021) concluded that the *Leucheria* species described as annual (and therein lumped into *L. tomentosa*) were actually perennial. Katinas2022 thus described the life history of this and all *Leucheria* species as perennial. But the presence of secondary vascular tissue is not per se an indicator of perenniality, and, furthermore, these authors' conclusions were based on study of herbarium material and not live plants in the field. Lavandero2024 referred to these species as annual, but did not mention/dispute the conclusions of Apodaca et al. (2021). For the record, I have studied plants of *L. tomentosa* (s. str.) on beaches near my current residence in El Quisco, Chile, and they are most definitely obligately semelparous annuals. The plants germinate following winter rains and desiccate/die completely in late spring to early summer following sexual reproduction. This is unlike certain facultatively annual perennial species in the exact same habitats, in which the caudex may survive summer drought and sprout again the following winter/spring (Hershkovitz, 2024f). I have found no early spring *L. tomentosa* plants that were holdovers from the previous growing season.

**(xi)** the new phylogenetic results invalidated macroevolutionary ecological conclusions of Pérez2020, because these were based on the erroneous phylogenetic reconstruction of Jara2017;

**(xii)** the compartmentalized capitula condition evident in almost all species of *Leucheria* sections *Leucheria* and *Cassiopea* (D.Don) Hook. (Katinas & Forte, 2020) is homologous with the pseudocephalic condition in *L.* sect. *Polyachyrus*, and it is most likely derived rather than ancestral in *Leucheria* (contra Katinas & Forte, 2020);

**(xiii)** the monotypic southern Patagonian herbaceous genus *Macrachaenium* Hook.f. pertains to Mutisieae rather than, as currently classified, Nassauvieae;

**(xiv)** the phylogenetic relations of the monotypic shrubby Chilean genus *Spinoliva* G.Sancho, Luebert & Katinas among Nassauvieae remain unresolved because of incongruence between the plastid and nuclear DNA trees;

**(xv)** plastid and nuclear DNA trees are incongruent at additional nodes in the *Leucheria* phylogeny; and

**(xvi)** the accuracy of the *Leucheria* and Nassauvieae ITS trees may be affected by nonstationarity of the molecular evolutionary process.

Lavandero2024 later described a new species of *Leucheria* and also, "*aiming to re-evaluate the phylogeny of* Leucheria *presented by Jara-Arancio et al. (2017)*," published a new phylogenetic analysis of this genus and Nassauvieae based on the same three loci used above. They (and their peer-reviewers and editor) evidently overlooked Hershkovitz (2024a–d).[7] Thus, they reported as novelty eight of the 16 conclusions listed above,[8] in particular regarding the relations of *Polyachyrus*,[9] *Oxyphyllum*, *Spinoliva*, and *Macrachaenium*, as well as polyphyly of genotypes of *L. tomentosa* sensu Katinas2022[10] and the incongruencies between plastid and nuclear DNA trees.[11]

For broader Nassauvieae, as in Hershkovitz (2024c), Lavandero2024 used published sequences, mainly those generated by Jara2017. But for *Leucheria*, they generated *new* sequences for 33 taxa[12] and did not use any of the published sequences generated for 46 taxa by Jara2017. This, they attributed to "*conflicts in the identification of herbarium specimens*" used in the latter. They gave no examples. They also omitted acknowledgement that they had used the Jara2017 *Leucheria* sequences in their earlier

---

[7] Hershkovitz (2024a–d) were published online 2–3 months before Lavandero2024 was *submitted* for publication.

[8] Conclusions **iii**–**vi**, **viii**, and **xiii**–**xv**.

[9] "*A* ***remarkable finding*** *is that* Leucheria*…appears as paraphyletic, as it includes* Polyachyrus…[**emphasis mine**]" (Lavandero2024: 320).

[10] "*…interestingly,* ***our*** *phylogenetic results indicated that* Leucheria tomentosa *(Less.) Crisci, as recognized by Katinas et al. (2022), is not monophyletic…***[emphasis mine]**" (Lavandero2024: 330). They did not report that the same result was obtained by Jara2017 using the *same* loci, or that, while Apodaca et al. (2021) and Katinas2022 lumped the annual species under *L. tomentosa*, they *acknowledged* their genotypic polyphyly per Jara2017. Pérez2020 (including Lavandero) published a "pruned" version of the Jara2017 tree, which *also* showed the annual species as polyphyletic.

[11] It must be emphasized that *all* of the data analyzed in Hershkovitz (2024a–c) except for sequences from two later-described species *were available to Jara2017*. Thus, Jara2017 – and later Pérez2020 and Lavandero2020 – *ought* to have found these same results. Lavandero2024 did not acknowledge that what they reported as discoveries were, effectively, corrections of their own earlier errors, oversights, and omissions. Likewise, one might (evidently unrealistically) expect that at least one person among the various reviewers and editors would have discovered at least one error/oversight/omission.

[12] 29 of which (nominally) were sampled by Jara2017.

phylogenetic analysis (Lavandero2020), or that they had used the Jara2017 phylogenetic tree for their macroevolutionary ecological analysis (Pérez2020; cf. Hershkovitz, 2024b).

But even if misidentified, the preemptive exclusion of *all* of the Jara2017 sequence data is peculiar and defies logic. The *primary* obstacle to the robustness of molecular phylogenetic conclusions at the interspecific level is lack of *sampling*, both *within* and among species. Thus, inclusion of the Jara2017 data, *at the very least* for species not sampled by Lavandero2024, would have been prudent from the standpoint of scientific rigor. Moreover, Lavandero2024 evidently examined the voucher specimens from Jara2017,[13] each including precise geographic provenance data. Lavandero2024 noted that some critical diagnostic characters are not preserved in herbarium specimens,[14] but, given their professed expertise on *Leucheria* taxonomy and evidently extensive field studies, it seems odd indeed that they could not identify *any* of the 46 Jara2017 vouchers, not even to "cf.," for the highly heuristic and desirable purpose of including the available DNA sequence data in their phylogenetic analysis. Finally, although they failed to mention it, Lavandero2024's combined data tree for eleven *L.* sect. *Lasiorrhiza* species is *congruent* with the Lavandero2020 tree generated using sixteen Jara2017 sequences and identifications.[15] Thus, their allegation of Jara2017 specimen misidentification is not only undocumented, it is *counterevidenced* by their *own* analysis using *different* specimens. Perhaps the sequences were excluded for some *other* reason that the authors declined to divulge.[16]

Besides excluding the Jara2017 data, Lavandero2024 is otherwise problematic. In particular, they indicated that they aligned the sequences for each locus using the alignment tool MAFFT (Katoh, 2002). But they did not specify the alignment options, did not indicate whether or not they edited the output, and they did not make available the alignment itself.[17] As I discussed in Hershkovitz (2024c), automated

---

[13] "*A systematic examination of herbarium specimens of* Leucheria *found at CONC and SGO was carried out*." (Lavandero2024: 317).

[14] This is a red herring, in any case. Lavandero2024 noted that the new species, *Leucheria peteroana* Lavandero, was most similar to *L. runcinata* D.Don, differing by some traits that might not be preserved in herbarium specimens. But these species are *highly* divergent at the examined DNA loci.

[15] Technically, it is congruent, but there is a subtle discrepancy. The relations of *L. hahnii* Franch are congruent in the two trees. But Katinas2022 included this species in *L. suaveolens* Druce, and this is how the Jara2017 and Lavandero2024 *L. hahnii* sequences are classified in GenBank. But the Jara2017 sequences *submitted* as *L. suaveolens* are phylogenetically divergent from the *L. hahnii* sequences, which I attribute to DNA contamination (discussed later). Since Lavandero2024 excluded the Jara2017 sequences, this discrepancy is not evident in their trees.

[16] When Lavandero2020 realigned/reanalyzed the Jara2017 sequences, they had available to them the Jara2017 alignment showing the excessive sequence lengths and poor alignability of the untrimmed chromatographic noise in the flanking regions. Hershkovitz (2024b) mused whether or not Lavandero2020 thusly had become aware of the inaccuracies of the Jara2017 sequences and alignment but, for political reasons, preferred to not report them. The only alternative explanation was lack of researcher competence and rigor. This reasoning extends to Lavandero2024's complete replacement of the Jara2017 sequences with new sequences from a smaller number of species. Note that they rejected the Jara2017 sequences because of identification "conflict," which specifies only a *disagreement* and not per se identification (or any other sort of) "error." Also, while Lavandero2024 summarily rejected the Jara2017 sequences, they did *not* summarily reject the Jara2017 *results*, nor the implications for the conclusions of Pérez2020 and Lavandero2020. One explanation is double-downed "cover up" of errors of Jara2017, Pérez2020 and Lavandero2020. This, of course, would be scientific misconduct. The other explanation logically implies that Lavandero2024 cannot identify even approximately well-documented *Leucheria* herbarium specimens, that they *still* are oblivious to the severe errors in the Jara2017 data and results, and that they cannot appreciate the implications of faulty Jara2017 data – for *whatever* reason – for the conclusions of Pérez2020 and Lavandero2020. This explanation implicates extraordinary incompetence. I am just framing the question scientifically here. I leave it to the readers to study the evidence and draw their own conclusions.

[17] For the broader Nassauvieae analysis, Lavandero2024 did use the Jara2017 sequences, which include untrimmed chromatographic noise in the flanking regions. They did not indicate whether or not the noisy regions were trimmed.

alignment of these loci, which include regions hypervariable for *both* sequence *and* length,[18] is problematic across this phylogenetic depth, viz. Nassauvieae and Mutisieae (Mutisioideae) and the Barnadesioideae outgroups. The hypervariable regions tend to align adequately among clusters of closely related taxa, but the clusters themselves will be superposed arbitrarily, resulting in spurious "substitutions" at spuriously aligned sites. Lavandero2024 did not mention these regions, nor did Lavandero2020.

In summary, Lavandero2024 independently corroborated several findings of Hershkovitz (2024a–c), but using varying degrees of different taxon sampling, data, alignment, and phylogenetic analytical methods. This can be viewed as a measure of robustness of the points of agreement, but only tentatively in the absence of the Lavandero2024 alignment. The work also provided new sequence data for 33 *Leucheria* taxa, which is useful given the poor quality of many of the sequences reported by Jara2017. At the same time, the wholesale exclusion of the Jara2017 sequences renders incomplete Lavandero2024's analysis of the *Leucheria* crown group. Besides this, Lavandero2024 was essentially a mechanical phylogenetic analysis, reporting mainly topological results with little critical discussion compared to that presented in Hershkovitz (2024a–c). Thus, the purpose of the present work is to consolidate the new and existing data into a revised analysis of Hershkovitz (2024b), thereby improving accuracy, precision, and documentation, and also shedding additional light on the taxonomy of *Leucheria*. In addition, comparative analysis of recent *Leucheria* research provides a backdrop for review of the mathematical nature of phylogenetic trees, data set incongruence, data support, operations of phylogenetic reconstructive methods, the relation between organisms and genomes, and relevant biological and evolutionary epistemology

## Materials and Methods

I obtained the *Leucheria* ITS, *trnL-trnF*, and *rpl32-trnL* sequences reported in Lavandero2024[19] for integration with the Hershkovitz (2024b) data (Appendix 1). I discussed in Hershkovitz (2024c) the properties of the sequences of these loci, in general and in these particular taxa. Preliminary to the analysis described below, using an online platform,[20] I evaluated MAFFT performance for the Hershkovitz (2024c) ITS and *rpl32-trnL* data. While the alignment was superior to that using CLUSTAL W (Thompson et al., 2002), hypervariable regions in both loci still aligned somewhat arbitrarily. In particular, as expected, alignment of regions hypervariable for *both* length and sequence was adequate among clusters of closely related (viz. similar) sequences, but still inadequate among clusters, resulting in spurious "substitutions" at sites that cannot be considered homologous. Thus, I edited the alignments manually.

The ITS alignment begins at the canonical ITS1 5' end and extends to the fifth base upstream of the canonical ITS2 3' end (cf. Hershkovitz & Zimmer, 1996). The GenBank sequence documents used here for the *trnL-trnF* and *rpl32-trnL* regions variously begin and end in different positions. I was able to construct a "consensus" alignment for each that maximized length and minimized missing data. For *trnL-trnF*, the alignment begins within the *trnL*(UAG) intron 37 bases upstream of the 3' end of the first exon and ends five bases downstream of the canonical 5' end of the *trnF*(GAA) gene. The *rpl32-trnL*

[18] For example, *rpl32-trnL* includes (sometimes adjacent) length-variable single-base repeats of the form $A_{(n)}$-$T_{(n)}$, as well as other simple repeats. These have been referred to as "hotspots" in other literature, e.g., Böhnert et al. (2019 [hereafter Böhnert2019]). The length variability owes to DNA replication "slippage" rather than base misincorporation (viz. "substitution").

[19] Plus one additional sequence reported elsewhere.

[20] https://www.genome.jp/tools-bin/mafft

alignment begins *within* the *rpl32* gene 88 bases upstream of the canonical 5' end of the intergenic spacer and ends 32 bases upstream of the canonical 5' end of the *trnL*(UAG) gene.

Following preliminary alignments, I manually screened the data to cull "significantly" errant or otherwise inadequate sequences besides those I culled in Hershkovitz (2024b). I considered a sequence to be errant if it contained multiple substitutions and/or indels at sites otherwise highly or completely conserved across the *Leucheria* crown group alignment. I focused especially on nominal species sampled twice, and I deleted the "most" errant sequence. This usually was the Jara2017 sequence, but in some cases it was the Lavandero2024 sequence. Still, I attempted to retain as many sequences as possible, so I did not discard a sequence it if contained only "plausible" substitutions and/or uninformative indels that were inconsequential to the phylogenetic analysis.

I also deleted from the current work an additional sequence included in the Hershkovitz (2024b) analysis, because I failed to notice its sequence errors. This was the ITS sequence for *L. diemii* Cabrera var. *diemii*. Based on morphology and DNA sequence divergence, Jara-Arancio et al. (2019) segregated *L. diemii* var. *purpurea* Ratto, M.Bello & Adr.Bartoli as a distinct species, *L. arancioi* Jara-Arancio, Ratto & Adr.Bartoli. Katinas2022 rejected their morphological evidence, and restored varietal status to this taxon.

Here, I reject also the DNA evidence. Jara-Arancio et al. (2019) reported that they found 12 ITS sites that differed between the varieties of *L. diemii*. I located these, plus two indels. Ten of the 12 "substitutions" and both indels occur in a 68-base stretch that extends from the 3' end of the 5.8S rDNA sequence to canonical base position 14 of the ITS2 sequence (Fig. 1). This sequence span is highly conserved across angiosperms (Hershkovitz & Lewis, 1996; Hershkovitz & Zimmer, 1996). In all cases, one of the varieties (usually *L. diemii* var. *diemii*) differs not only from the other, but from *all* of the *Leucheria* crown group sequences and all examined Nassauvieae sequences. In other words, the polymorphisms reported by Jara-Arancio et al. (2019) most likely are *sequencing artifacts*. In most cases, they appear to be chromatographic "smears," because the differences are repeats of the adjacent base (Hershkovitz, 2024b). I did not bother to analyze the sequence differences that Jara-Arancio et al. (2019) reported for other species pairs. Based on my analysis of Jara2017 sequences/alignments (Hershkovitz, 2024b), I have no doubt that these also are errors.

For the present analysis, I created two alignments (see Supplemental Files) one for ITS and the other for cpDNA, combining the *trnL-trnF* and *rpl32-trnL* sequences. But I included in the latter only taxon samples for which at least an *rpl32-trnL* sequence was available and none for which only a *trnL-trnF* sequence was available. This is because ca. 75% of the informative variation at these combined loci owes to *rpl32-trnL* (see Results). Thus, including a taxon with only a *trnL-trnF* sequence would cause the taxon sample to associate spuriously in the bootstrap analysis, because many/most replicates would not sample the few informative sites. BPs across the entire tree would be reduced. Thus, the cpDNA data set included both of the sequences for all but one taxon sample, for which only *rpl32-trnL* was available.

The ITS data set included 60 taxon samples, of which 56 pertained to *Leucheria* sensu Katinas2022, etc. The cpDNA data set included 72 taxon samples, 69 of which pertained to *Leucheria* sensu Katinas2022, etc. Several sequences probably still include artifacts, especially spurious single-base indels. As long as these did not yield obvious artifacts in the phylogenetic analyses, these sequences were retained in the data unadulterated. For computational efficiency, I condensed the data further into "genotypes:" combining sequences that were identical for scored characters (but ignoring spurious uninformative indels). As a result, the ITS data includes 45 unique sequences and the cpDNA data 47 (46 in the analysis excluding scored indel characters). The positions of the samples with identical sequences are indicated in the tree illustrations.

All three loci, but especially the *rpl32-trnL* locus, manifested length heterogeneity owing to indels. Ignoring incomplete sequences, the lengths of the aligned portions of the ITS, *trnL-trnF*, and *rpl32-trnL* sequences range from, respectively, ca. 636–644bp, 806–840bp, and 660–946bp.[21] ML and related statistical analyses generally ignore parsimony-informative indels, but they can be exploited in MP analysis. I scored informative indels for up to four states (plus N for "ambiguous"). But hypervariable and ambiguously aligned indels, e.g., length-variable simple repeat regions, were ignored. These indels were determined to add too much homoplasious "noise" to the analysis (Hershkovitz, 2024c). For ITS, *trnL-trnF*, and *rpl32-trnL*, respectively, two, four, and eleven informative indels were scored.

As in Hershkovitz (2024a), I performed maximum parsimony (MP), frequentist maximum likelihood (ML), and MP bootstrap (500 replicates) analyses for the separate data sets using PAUP version 4 (Swofford, 2003). The ML analyses applied a parameterized first-order reversible[22] Markov (viz. stochastic or indeterminate) model of DNA sequence evolution. The parameterized substitution model was chosen by ModelTest using the data and an MP tree under the AICc criterion (Posada & Buckley, 2004) as implemented in PAUP. For ITS, the model selected was the symmetrical model[23] with equal base frequencies. But as I noted in Hershkovitz (2024a), %GC among MP-informative sites varies markedly (45–72%). This means that the substitution dynamics have not been stationary,[24] and that the "true" substitution model is *not* reversible[25] (see Hershkovitz, 2021a). This, paradoxically, possibly explains why ModelTest preferred an equal base frequency over an empirically estimated model, since no *particular* base frequency specification improved the likelihood. For the cpDNA data, the model selected was the transversion model (transition rates equal) with estimated base frequencies and gamma.

Hershkovitz (2024c) *erroneously* reported that the %GC among MP-informative sites is uniformly low in Nassauvieae cpDNA sequences. The figures reported represent all sites, not MP-informative sites. In fact, the MP-informative sites range from slightly AT-rich to markedly GC-rich (see Results). Nonetheless, in the present analysis, unlike ITS, ModelTest selected a model[26] that included an estimated base frequency parameter, and the frequencies approximated those of the whole sequences. This can be explained in terms of the small number of variable/informative sites in the cpDNA alignment. For ITS, 26% of the aligned sites are MP-informative. For the cpDNA alignment, the figure is only 4%. Model selection for the base frequency parameter thus may be an artifact of data set construction. For example, if the rDNA analysis had included the entire rDNA 18–26S cistron, some 4kb, ModelTest probably would have selected estimated rather than equal base frequencies. In this case, ca. 90% of the cistron sites would

---

[21] These are approximations, given that a few single-base indels in each data set may be spurious.

[22] This means that the substitution rates in both directions are equal.

[23] SYM ["abcde;" all rates different] plus gamma model: AC = 1.0192849, AG = 2.9037188, AT = 1.0691928, CG = 0.30931535, CT = 8.8027107, GT = 1; base frequencies equal (pA,C,G,T = 0.25); gamma shape = 0.42141108. Note that the precision is excessive given the likely variance, though not as excessive as for the cpDNA data. Rounding to a single decimal would yield a "abade" model, and this probably would yield the same phylogenetic results.

[24] Model nonstationarity is the same as model instability or *hyperstochasticity*. Gorban (2017) monographed this phenomenon generally, though not with specific reference to molecular evolution.

[25] However, hyperstochastic substitution models, while not statistically completely intractable, are computationally impractical, partially because of non-nestedness and partially because of operational small sample sizes. Substitution sample sizes for conventional molecular phylogenetic loci already limit precision of even stationary models. Phylogenomic analyses may be a different matter. Other sorts of evolutionary analyses have added hyperstochastic "jump" parameters to accommodate nonstationarity, viz. abrupt transitions between two different rates.

[26] TVM ["abcdbe;" one transition rate, four transversion rates] plus gamma model: AC = 0.91168947, AG = CT = 0.70848648, AT = 0.27555982, CG = 0.82789942, GT = 1); base frequencies: pA = 0.353866, pC = 0.15408, pG = 0.153527, pT = 0.338527; gamma shape = 0.51413. Note that the precision is absurd given the variance for the small number of variable sites. Rounding to a single decimal would yield close to an "abcabe" model, and this probably would yield the same phylogenetic results.

be invariant and GC-rich.[27] The *proportion* of informative sites would be much smaller, hence the degree of *apparent* base frequency variation among sequences would be much less.

Because of incongruencies between the ITS and cpDNA trees, and also because the sequence sampling in the data sets differed, I did not perform a combined data analysis. I also did not perform [pseudo-]Bayesian ML analysis, because I consider this method to be "induction on steroids" (Hershkovitz, 2021a).[28] The posterior probability (PP) values commonly are misinterpreted as data support (e.g., by Swenson et al., 2025; cf. Hershkovitz, 2021a). I have conjectured that they are a biased estimate of the probability (or confidence interval?[29]) that the corresponding branch exists in the "true" ML tree given the *inherently* inaccurate substitution model estimate.

## Results

Results of the phylogenetic analyses obtained here are illustrated in Figures 2–4 for ITS and Figures 5–6 for cpDNA.

### 1. Comparison of ITS and cpDNA data and trees

**i. Amount and distribution of evolutionary change.** The ITS, *trnL-trnF*, and *rpl32-trnL* data (without indel characters) include, respectively, 170, 20, and 52 MP-informative sites over alignments of, respectively, 657, 850, and 977 sites. Informative indels added, respectively, 2, 4, and 11 characters. These are fewer than scored in Hershkovitz (2024a), where additional scored indels were found to be excessively homoplasious (see Hershkovitz, 2024b, c).

The pattern and distribution of phylogenetic substitutions are illustrated in the ITS and cpDNA ML phylograms (Figures 4, 6). The ML trees are scaled to exactly an order of magnitude difference in terms

[27] Even within the ITS2 sequence, regions *relatively* more conserved among angiosperms are slightly to markedly GC-rich (Hershkovitz & Zimmer, 1996).

[28] Technically, what is termed Bayesian estimation in ecological and evolutionary biological analysis is a [pseudo-]Bayesian ML optimization algorithm rather than a Bayesian statistical analysis per se. The method is called "Bayesian" simply because Bayes' formula is incorporated into the algorithm at a particular optimization step. An example of true Bayesian analysis is evaluating the probability that a woman carries the X-linked hemophilia allele given that her brother is hemophiliac. The analytical probability is 0.5 and is *static* with successive births of healthy sons. The Bayesian *prior* probability is 0.5, but the *posterior* probability (PP) *decreases* towards zero with the successive birth of each healthy son. Here, each birth is a *new* statistically *independent and identically distributed* (IID) empirical observation whose outcome reflects the *cumulative* probability distribution and rather than the a priori analytical probability. In both cases, the probability becomes 1.0 when a hemophiliac son is born. In [pseudo-]Bayesian ML optimization, the PPs are based on a distribution not of new IID observations but of biased pseudo-observations (usually) generated by a Metropolis-coupled Monte Carlo (MCMC) algorithm. The pseudo-observations are generated by constrained (viz. nonindependent, hence biased) branch/parameter modifications from "current" calculations, and the ML is calculated from the *same* (not new) observations. The MCMC algorithm, using Bayes' Theorem and a *specified* acceptance criterion, nonetheless "accepts" some *suboptimal* moves as *provisional* optima for purposes of continuing the search for a *global* optimum. This is how this method (supposedly) escapes the "local minimum" trap. Bromham et al.'s (2018) comprehensible description should render clear that it is a (biased) ML optimization algorithm rather than a Bayesian statistical analysis s. str. (cf. Hershkovitz, 2021a; see also Results). Conventional frequentist ML optimization, such as in PAUP and RAxML, applies a hill-climbing algorithm that *never* prefers a less optimal solution in favor of a more optimal one. The "local minimum" trap is avoided a different way, viz. by branch swapping efficiency and, if desired, *optional* retention of certain suboptimal trees for further algorithmic exploration.

[29] I will not elaborate here the technical reasons for this possible ambiguity.

of "substitutions per site."[30] Thus, it can be "eyeballed" that the number of substitutions along the ITS tree is somewhat less than ten times more than that of the cpDNA tree. This is what the MP phylograms (not shown) indicate, although the distribution of change among internal and external branches differ. This is because MP minimizes the number of changes along a branch, while ML "corrects" for the probability of "unseen" homoplasious changes that occurred along that branch with increasing empirically observed divergence.[31]

Although ITS seems far more "informative" than the cpDNA data, it also is more "misinformative," viz. more homoplasious. The rescaled consistency indices for ITS and cpDNA are, respectively, 0.56 and 0.89, the latter 0.87 with indel characters included. Still, total "resolution" appears slightly greater for the ITS data: 18/43 nodes are resolved with 70+ BP. (Fig. 2) versus 17/46 nodes for the cpDNA data. But notably, neither data set resolves many terminal interspecific bifurcations with this level of support: only six for ITS and three for cpDNA. However, these figures would be greater if, e.g., each terminal taxon were duplicated. Thus, a simple "resolution index" value of a data set partially is an artifact of data set construction.

---

[30] Substitutions/site is an *abstract* measurement, taking into account the length of the alignment. Here, the cpDNA alignment is ca. 2.5X longer than the ITS alignment, while the number of MP-informative sites in ITS is ca. 2.5X more than the cpDNA data. This means that, with minimal homoplasy, the amount of total change of ITS along the *internal* branches of the tree should be ca. 6 times that of cpDNA. And this is approximately what is evident in the ML phylograms. But ML arrives at this approximation in a different way, selecting the tree and branch lengths that optimize the *likelihood* of the observed data according to, in this case, a *presumed* stochastic base substitution model of the probability of (a particular) *instantaneous* change at *any* site. Hence, the branch lengths are scaled not according to the number of observed base differences, but rather the ML calculation of substitutions *per aligned site* across *all* sites. This is the summed *probability* of sitewise change that best explains the observed differences given the presumed statistical difference-generating process (viz. the substitution model). However, it must be emphasized that ML and related statistical phylogenetic methods are framed in statistical and not *biological* theory: "*…the complex mutational processes producing real sequence diversity are never fully captured by nucleotide substitution models*" (Schwartz & Mueller, 2010: 16); "*…assumptions in phylogenetics are made for statistical tractability rather than being necessarily based on understanding of the underlying evolutionary processes*" (Bromham et al., 2018: 3). For example, Cao et al. (2022) described a procedure for calculating ML-based evolutionary distances among ITS2 sequences that corrects for compensatory substitutions that maintain RNA secondary structure base-pairing. While perhaps more realistic in this way than conventional sitewise models, the method has other shortcomings. Since any and all statistical substitution models are not "real," Hershkovitz (2021a) termed them "fake" and, for other epistemological reasons, described molecular and indeed all evolutionary processes as *idiosyncratic*.

[31] ML and related statistical phylogenetic methods are quantitative and inductive, but their justification is rooted in qualitative reasoning and abduction. ML attempts to "probabilisticize" a posteriori logical *reasoning* abduced from *observations* of patterns in aligned DNA sequence base variation. ML further takes into account the empirical observation that, unlike the case for complex morphological characters, there is no *physical* way to distinguish between homoplasy and synapomorphy, such that, as sequences diverge, more homoplasious changes occur, until the pairwise sequence differences are physically and statistically indistinguishable from that of a pair of random sequences. As Swofford et al. (1996) noted, MP also can be cast in terms of an ML model in which all sites in an alignment evolve *completely* independently of any process that might affect all sites *uniformly*, and it ignores the statistical homoplasy/divergence relation. Thus, while the ML method yields solutions that are not "maximally parsimonious" per MP, it actually bases upon more *parsimonious* explanations for observed patterns, and it incorporates these explanations into the ML model. This clarified, the question of ML model *adequacy* in practice remains problematic. One model/tree might fit the data better than another, but this does not mean that the model/tree is *correct* (cf. Bromham et al., 2018). Moreover, statistical precision is decreased with increasing parameterization, which *effectively* decreases the quantity of IID observations per parameter, hence increases variance. In practice, as parameterization reduces the number of effective observations, simpler models may "perform" better than a "true" complex one (see Hershkovitz, 2021a).

Ignoring topological incongruencies (described below), it can be seen from the ML phylograms that, independently of topology, change in the two data sets is not distributed *proportionally*. For example, the *L.* sect. *Lasiorrhiza* stem branch in the ITS tree is slightly more than twice as long as the same branch in the cpDNA tree. Given the scaling, it should be no more than 1.5X as along. And note also the differential extension of the *L.* sect. *Lasiorrhiza* crown node *relative* to those of the other sections. Note also a long branch in the cpDNA tree subtending a subclade of *L.* sect. *Cassiopea*. A topologically similar branch is present in the ITS tree, but it is disproportionally short and, while it has 100% BP in the cpDNA MP analysis (Fig. 5), it only has 44% BP in the ITS MP analysis (Fig. 2; the "without indel" support). As a final example, note that taxon samples identical to *L. coerulescens* J.Rémy in the ITS ML phylogram are *different* in the cpDNA ML phylogram. Yet, ITS divergence is, on average, ca. 6X that of the cpDNA sequences. If molecular evolution in the two genomes was proportional within a lineage, the branch lengths should be proportional all across the tree. The significance of these observations will be discussed later.

**ii. Base compositional biases.** As reported earlier (Hershkovitz, 2024a, c), the ITS sequences vary considerably in %GC at variable, in particular MP-informative sites. The differences appear to be phylogenetically patterned, as illustrated in Fig. 2. The outgroup and *L.* sect. *Leucheria* sequences are moderately GC-rich, those of *L.* sect. *Cassiopea* and *L.* sect. *Polyachyrus* more so, and those of *L.* sect. *Lasiorrhiza* actually slightly to moderately AT-rich. *Oxyphyllum* is extremely GC-rich. Hershkovitz (2024a) suggested that the %GC of *Oxyphyllum* might cause it to spuriously attract to other GC-rich taxa in MP analysis. To appreciate this, I constructed a dataset comprising 17 ITS sequences from *Oxyphyllum*, *L. floribunda* DC, and the *L.* sect. *Lasiorrhiza* samples. The number of variable sites[32] was ±127, whose %GC was, respectively, 77%, 66%, and a mean of 41%. From this, it can be calculated that at 98 G/C sites of *Oxyphyllum*, ca. 46 of these are A/T in *L.* sect. *Lasiorrhiza* sequences. The differences are mostly C-T transitions, followed by G-A transitions. Meanwhile, the G/C-A/T differences between *Oxyphyllum* and *L. floribunda* occur at only 14 sites. Thus, the strong base bias should affect MP results, and indeed, the MP trees show *Oxyphyllum* as sister or neighbor to *L. floribunda*.

These results *alone* do not implicate spurious branch attraction, since high %GC may be a *true* "synapomorphy" of these taxa. Indeed, similarity of %GC *within* sections indicates that it is phylogenetically conserved at least at the *interspecific* level. The evidence for spurious branch attraction emerges in the ITS ML phylogram (Fig. 4), where it can be seen that, despite its extremely high %GC, *Oxyphyllum* is sister to *L.* sect. *Lasiorrhiza*, which is AT-rich. Notably, the ML model itself presumed equal base frequencies (see above), hence could not have per se "corrected" for base bias, at least not directly. The model actually corrected for extremely high transition rates, especially C-T transitions, and also for among-site rate heterogeneity. Applying these corrections ("weights" sensu Williams & Ebach, 2020: Ch. 8), there is nothing else in the ITS besides %GC that suggests an especially close relationship between *Oxyphyllum* and *L. floribunda*. At the same time, a propos the long branch subtending the *L.* sect. *Lasiorrhiza* crown node in Fig. 4, the same applies. This branch length already is at least partially but *indirectly* corrected for base compositional bias via correction for the high C-T homoplasy and among-site rate heterogeneity.

Does base compositional bias affect MP results for the cpDNA? Hershkovitz (2024c) erroneously reported that %GC at variable sites in Nassauvieae cpDNA sequences was uniformly low. But the figures reported there were for the entire sequence, not the variable sites. Figure 5 shows the values for MP-informative sites. These range from slightly AT-rich to markedly GC-rich. Interestingly, the overall *range* of GC-richness in the cpDNA MP-informative sites is essentially identical to that in ITS if extreme values of *Oxyphyllum* and *L. floribunda* are excluded. But there is not a phylogenetic correlation between the two loci. For ITS, AT-richness occurs in *L.* sect. *Lasiorrhiza*, while moderate GC-richness characterizes

[32] Here I use variable sites, because these may be MP-informative in the full data set.

*L.* sect. *Leucheria*. These values are more or less reversed in the cpDNA data. Both the ITS and cpDNA sequences are GC-rich in *L.* sect. *Cassiopea.* But this section includes two cpDNA modes: 57% GC or 60–63% GC. The modes *appear* to have some phylogenetic correlation in the MP bootstrap excluding indel characters, but it is not absolute. The ML topology is compatible with the MP bootstrap excluding indels, so there is no evidence that base compositional bias does or does not cause spurious branch attraction in the latter. But it is important to note that the number of MP-informative characters is only 72. Thus, the difference between the extremes of the entire data set, 44–63% GC, involves no more than 13 sites and, within sections, fewer.

**iii. The effect of indel characters.** Hershkovitz (2024a, cf. 2024b) found that inclusion of indel characters in MP analysis of this group increased homoplasy and reduced resolution, especially for the cpDNA data. This was because that work attempted to score single-base repeat polymorphisms as discrete indels. Thus, indel characters were not included in Hershkovitz (2024b), which is equivalent in *scope* to the present analysis, but used only the Jara2017 data. Hershkovitz (2024c) incorporated only unambiguous indel characters in a broader analysis of Nassauvieae. The effect on resolution was small and lineage-specific, but indels increased MP BP for the transfer of *Macrachaenium* Hook.f. from Nassauvieae to Mutisieae. Lavandero2024 later analyzed the same relationship using an ML bootstrap (viz., substitutions only) and reported BPs for the separate and combined data sets comparable to those of the Hershkovitz (2024c) MP bootstrap without indels.

In the present analysis, I scored indels more conservatively, yielding fewer characters than in Hershkovitz (2024c). The effect on MP BP was minimal,[33] else lineage-specific. An interesting example involves the relations of *L.* sect. *Polyachyrus* based on the ITS data. The relations are somewhat different than for the cpDNA data (Hershkovitz, 2024a, b; Lavandero2024). The latter strongly support *L.* sect. *Polyachyrus* as sister to *L.* sect. *Cassiopea*. The former cluster *L.* sect. *Polyachyrus* with *L.* sect. *Macrobotrys*. This appears in the MP consensus (not bootstrap consensus), but without significant BP support for this relation, nor for the relation of this clade to other sections. As it happens, one of the ITS indel characters markedly increases BP for the relation supported by the cpDNA data. This level of support *still* is insignificant, and it remains less than the support for relations *otherwise* supported by ITS. But at least it suggests that the incongruence between the ITS and cpDNA is less than it appears (see also below).

As noted, inclusion of indel characters in the cpDNA analysis increases homoplasy. Its effects on BPs are mostly insignificant, except at two nodes highlighted in Fig. 5. Inclusion of indel characters markedly increases support for a subclade within *L.* sect. *Cassiopea*, and also for the nesting of *Oxyphyllum* within *Leucheria*, though support for the latter still is insignificant. But review of the alignment found the cause of the BP differences: a 19bp tandem duplication within the *trnL-trnF* locus (Fig. 7). The duplication is absent in *Marticorenia* (the outgroup) and *L.* sect. *Lasiorrhiza*, but it is polymorphic within both *L.* sections *Leucheria* and *Cassiopea*, and it is present in *L.* sect. *Polyachyrus* and *Oxyphyllum*. There is no escaping the conclusion that the 19bp tandem duplication originated more than once. Nesting of *Oxyphyllum* within *Leucheria* reduces the minimum number of origins from three to two. This nesting evidently is at least *compatible* with the cpDNA substitution data. But this is no surprise, because, just as for the ITS data, the cpDNA substitution data alone does not resolve the relations of *Oxyphyllum* one way or another (Fig. 5: BPs without indel characters; cf. Fig. 6; cf. Figures 2–4).

[33] For the present data sets, I found that, for bootstraps with 500 replicates, BPs from replicate analyses commonly differed by 1–2% and sometimes up to 4%. I really should have done 1000 replicates.

## 2. Specimen identification assessment

As noted above, Lavandero2024 cited specimen misidentification as the reason for excluding the Jara2017 data from their analysis. Inclusion of the data yields numerous insights in the present analysis, among these, on the conflicts of specimen identification. Taxon acronyms in Figures 2–6 are flagged for source (Jara2017 or Lavandero2024), and several of the Jara2017 samples are enclosed in quotation marks, indicating "identification issues."

But this analysis is not as straightforward as one might hope. First, there does not yet exist the robust taxonomy of *Leucheria* presumed by such analysis (Lavandero2024). Second, even if there did, I do not have available to me the vouchers of either Jara2017 or Lavandero2024. But at least Jara2017 published the geographic coordinates of their vouchers, and this is useful for comparing with published species distributions (Crisci, 1976; Katinas2022) and online collections and specimen images (GBIF Secretariat, 2017; iNaturalist.org[34]). Besides this, there is not a 1:1 relation between DNA sequence genotype and species identity (discussed later). And even if there was, as I have noted, some of the DNA sequences evidently contain numerous errors and artifacts.

The above notwithstanding, the combined data analyses of Jara2017 and Lavandero2024 reveals that specimen misidentification is the least of the problems with the former. Lavandero2024 included 29 new samples of *Leucheria* species also sampled by Jara2017. As noted, the 29 new samples included eleven for *L.* sect. *Lasiorrhiza*, and the Lavandero2024 combined data tree for these species is congruent with the Lavandero2020 tree generated using the Jara2017 sequences and identifications.[35] The present analysis suggests "suspicious" identifications for only at most seven of the Jara2017 samples, and these problems are largely resolvable in view of the molecular data. Here, I judge that four of these Jara2017 samples were misidentified, one of which I discovered in Hershkovitz (2024b). Thus, there seems to be no a priori reason for Lavandero2024 to have *preemptively* excluded *all* of the Jara2017 sequences from their analysis…at least for the reason they claimed. Moreover, unlike me, Lavandero2024 had, besides the sequence data, and besides the luxury of a salary and public research funds, also the luxury of being able to examine all Chilean collections of *Leucheria*, including the Jara2017 vouchers. The fact that they *effectively* claimed to have been unable to identify any of the Jara2017 vouchers is problematic, to say the least. In any case, I describe the problematic Jara2017 samples below.

**i. "*L. coerulescens.*"** Hershkovitz (2024b) found that the ITS sequences of the Jara2017 "*L. coerulescens*" and *L. floribunda* samples were identical, and that the cpDNA sequences were similar and sister in the phylograms. The same ITS result was reproduced here. The Lavandero2024 *L. floribunda* ITS sequence is highly similar, but it was excluded from the analysis because the few differences appeared to be sequencing artifacts. Meanwhile, the Jara2017 cpDNA sequences for both species were excluded from the present analysis, because the *rpl32-trnL* sequences of both contained sequencing artifacts (including a peculiar 184 bp deletion in the "*L. coerulescens*" sequence). However, cpDNA sequences of both Jara2017 samples are highly similar to the Lavandero2024 *L. floribunda* sample. But the Lavandero2020 *L. coerulescens* ITS and cpDNA sequences map to *L.* sect. *Leucheria* and are most similar to Jara2017 *L. amoena* Phil. samples (Figures 4, 6). This is consistent with the morphology of these species, especially since *L. floribunda* is morphologically notably distinct from all other *Leucheria* species (Katinas2022; Lavandero2024). Thus, the Jara2017 "*L. coerulescens*" sequences correspond to *L. floribunda*. I cannot

[34] www.inaturalist.org. Lavandero is listed as the curator of the iNaturalist.org *Leucheria* images, although evidently not all species accepted by Lavandero are recognized by iNaturalist.org taxonomy.

[35] There is a subtle discrepancy involving *L. hahnii* and *L. suaveolens*, since Katinas2022 considered these to be taxonomic synonyms, whereas the Jara2017 sequences submitted under these names are phylogenetically divergent (see later discussion). Lavandero2024 did not notice this, because they only analyzed a sample that they identified as *L. hahnii*, whose sequences are essentially identical to the Jara2017 *L. hahnii* sequences.

judge whether the discrepancy owes to specimen misidentification or DNA sample mix-up/contamination. As for most Jara2017 sequences, the "*L. coerulescens*" DNA sample was extracted from a herbarium specimen. Hershkovitz (2024b) noted several examples of evidently low quality DNA and contamination.

**ii. "*L. congesta.*"** As noted in Hershkovitz (2024b), the Jara2017 cpDNA sequences reported for *L. congesta* D.Don are identical with those of *L. bridgesii* Hook. & Arn. as identified in Jara2017 (Fig. 6). The Jara2017 ITS sequence, however, was contaminated with *L.* sect. *Lasiorrhiza* ITS sequence. But the uncontaminated portion mapped to *L.* sect. *Cassiopea*. Katinas2022 included *L. congesta* in *L. runcinata*, which pertains to *L.* sect. *Leucheria*. All three species are sympatric. Based on discussion in Katinas2022, I judged that the *L. congesta* of Jara2017 is *L. bridgesii*. The ITS ML tree (Fig. 4) shows the Jara2017 and Lavandero2024 *L. bridgesii* genotypes as polyphyletic, but this may reflect true polymorphism and/or sequence error. The similarity of the Jara2017 "*L. congesta*" cpDNA sequence to the Lavandero2024 *L. bridgesii* sample is evidenced in Fig. 6.

**iii. *L. hieracioides*.** The Jara2017 and Lavandero2024 *L. hieracioides* Cass. samples are divergent in the cpDNA trees (Figures 5–6). This is partially because the latter and all species of its *L.* sect. *Leucheria* subclade share a 6bp deletion in the *trnL-trnF* sequence (Fig. 7). The Jara2017 *L. hieracioides* sample and the other *L.* sect. *Leucheria* subclade species lack this deletion. Hence, this sample occupies a position intermediate between these subclades in Figures 5–6. The Jara2017 *L. hieracioides* ITS sequence was excluded from the analysis because of apparent sequencing artifacts. But preliminary analyses indicated that it indeed clusters among the same species as the Lavandero2024 sequence, viz. the Jara2017 *L. hieracioides* ITS sequence *does not* manifest divergence similar to the Jara2017 cpDNA sequence. Note that I did not put the Jara2017 sample in quotes, because the present data do not permit identification of the "real" *L. hieracioides*. Multiple explanations are possible for the divergent Jara2017 cpDNA sequence, so no conclusions are drawn here.

**iv. "*L. oligocephala*."** For both ITS and cpDNA, the Lavandero2024 *L. oligocephala* J.Rémy sample maps to *L.* sect. *Cassiopea*,[36] while the Jara2017 sample maps to *L.* sect. *Leucheria* and is essentially identical to the *L. glandulosa* D.Don sequences (Figures 4, 6). Katinas2022 included both species in *L. tomentosa* (Less.) Crisci (see also below). Even though genotypes of the latter sensu Katinas2022 are polyphyletic, the Jara2017 identification underscores the apparent difficulty of discriminating between individuals from the two different sections.[37] Nonetheless, it appears that the Jara2017 "*L. oligocephala*" sample is misidentified.

**v. "*L. tenuis*."** The cpDNA sequences of the Jara2017 specimen identified as *L. tenuis* Less. are identical to those of the Lavandero2024 specimen identified as *L. glabriuscula* (Phil.) Reiche. The sequences map to *L.* sect. *Leucheria* and are sister to *L. glandulosa* (Fig. 6). Lavandero2024 did not obtain an ITS sequence for *L. glabriuscula*. But since the Jara2017 "*L. tenuis*" ITS sequence also is sister to *L. glandulosa* (Fig. 4), quite likely it corresponds to *L. glabriuscula*. Meanwhile, the Lavandero2024 *L. tenuis* ITS and cpDNA samples map to *L.* sect. *Cassiopea* in a position near *L. bridgesii* (Figures 4, 6). This is another example demonstrating the difficulty of distinguishing annual plants in these divergent sections and why Katinas2022 considered them all to be the same species. This example, however, is particularly complex, because the Type of *L. tenuis* is a Carlo Bertero collection supposedly from the Andean precordillera near Rancagua, Chile (Katinas2022). Bertero later collected what he identified as the same species from the coastal ranges at Quillota, Chile. Katinas2022 cited the Quillota specimens and

[36] The Lavandero2024 *L. oligocephala* ITS sequence is identical to that of the Lavandero2024 *L. tenuis* sample (Fig. 4; see below), whereas this identity is not evident in the cpDNA sequences (Fig. 6). This is peculiar, because the ITS sequences contain overall 2–3X more variation than the cpDNA sequences.

[37] The Lavandero2024 *L. tomentosa* sequence was excluded from the present analysis because of possible sequencing artifacts, but, like the Jara2017 *L. tomentosa* sample, it does pertain to *L.* sect. *Cassiopea*.

additional specimens from a *different* Rancagua locality – all having the *same* numbers – as syntypes.[38] The complication is that Hershkovitz (2020) determined that Bertero and/or whoever curated his collections after his death combined specimens from disparate localities (and years) and assigned to them the same apparent "collection" number. Hershkovitz (2020) deduced that Bertero's numbers were intended to correspond to species rather than collections, and that Bertero himself intended to consolidate later numbered synspecific collections into the earlier ones. Individuals with the same number but different provenances later were distributed as separate specimens, often with either multiple or erroneous localities and/or collection dates.[39] Thus, the provenance of specimens is not certain, and, in some cases, they may differ from that indicated on the specimen label. The significance here is that the Types of the various species that Katinas2022 classified in *L. tomentosa* are either precordilleran or coastal, and, in general, their individual ranges have been so-defined (Crisci, 1976). When recognized as distinct from *L. tomentosa*, *L. tenuis* is considered to be a precordilleran species. But there is a chance that the lectotype is from the coastal ranges. I presume that the Lavandero2024 collection is from the precordillera. If so, this demonstrates at least that annual species from both sections co-occur in the precordillera (contra Hershkovitz, 2024e). Whether species of both sections also co-occur in coastal regions is not clear. It depends, partially, on the taxonomic identity of the Quillota collections that Bertero identified as *L. tenuis*.

**vi. *L. salinae salinae*.**[40] The ITS and cpDNA MP bootstrap trees significantly differ with respect to the relations of the Jara2017 sample of *L. salinae* (R.Rémy) Dusén subsp. *salinae* (Figures 2, 5). They map to different otherwise well-supported subclades of *L.* sect. *Lasiorrhiza*. But in both trees, this sample localizes near the sample of *L. cantillanensis* Lavandero, viz. the ITS and cpDNA trees differ for *both* taxa. Adding to the mystery, the ITS and cpDNA trees also differ with respect to the position of the Lavandero2024 sample of *L. salinae*. In the ITS tree, this sample maps to the same subclade as the Jara2017 L. *salinae* subsp. *salinae* sample, but nonetheless manifests different relations therein. Meanwhile, cpDNA sequences for both *L. salinae* samples are identical and map to the other subclade. Thus, I suspect that the Jara2017 sample was correctly identified as *L. salina*, but its ITS sequence is either divergent or contains errors or both. I discuss this matter further below.

**vii. "*L. suaveolens*."** Jara2017, Lavandero2020, and Hershkovitz (2024b) found that for both ITS and cpDNA, the Jara2017 sample of *L. suaveolens* was sister to a clade comprising the Central Andes species *L. daucifolia* (D.Don) Crisci and *L. pteropogon* (Griseb.) Cabrera (see also below). But the Jara2017 "*L. suaveolens*" ITS sequence was only partially recoverable, evidently consequent to low

---

[38] All cited Type material cited by Katinas2022 indicate "*C. Bertero 159*." The collection date of the designated type indicates October, 1828. The Quillota syntype collection indicates October, 1829. The second Rancagua locality syntype collection indicates October, 1829, but Hershkovitz (2020) demonstrated that all Bertero collections from near Rancagua were collected in 1828.

[39] *This* is why the *L. tomentosa* syntype specimen from a nontype Rancagua locality indicates 1829: this is the date of a later Quillota collection transposed to a specimen collected earlier near Rancagua. More importantly, this demonstrates that Bertero collections are labeled with mixed locality/date data, hence the actual provenance/date of the material cannot be established. The possible exceptions are specimens in TO-HG bearing Bertero's original labels and described by Colla (see Hershkovitz, 2020). This material was sent to Colla after Bertero's disappearance but before remaining Bertero collections were hopelessly mixed up during distribution. We can believe (but not know) that the TO-HG specimens and labels were not mixed/adulterated. In this case, while the locality of the *L. tenuis* lectotype designated by Apodaca et al. (2021) indeed is that stated by Colla, the lectotype itself is a P specimen. Its actual provenance and date therefore can be considered as "uncertain." The same is true for three additional names lectotypified by Apodaca et al. (2021; cf. Katinas 2022) with Bertero collections in P: *L. abbreviata* Steud. (as "*Leuceria*"), *L. senecioides* Hook. & Arn. var. *purpurascens* DC, and *Chabraea* [= *Leucheria*] *elongata* Bertero ex Colla. Misciting Delprete et al. (2002), Katinas2022 reported incorrectly that no Bertero collections have been found in TO-HG (cf. Hershkovitz, 2020, 2023).

[40] Katinas2022 pointed out that the correct spelling for this epithet is "*salinae*" rather than "*salina*."

quality genomic DNA extracted from a herbarium specimen (Hershkovitz, 2024b). Thus, this ITS sequence was not included in the present analysis. But what I did not appreciate in Hershkovitz (2024b) was the peculiarity of a major "geographic disjunction." The Type of *L. suaveolens* is from the Falkland Islands (Islas Malvinas; Katinas2022). Katinas2022 included in this species *L. hahnii*, whose Type is from southernmost Patagonia. For both ITS and cpDNA, the Jara2017 *L. hahnii* specimen associates, as expected, with species of the more southerly subclade of *L.* sect. *Lasiorrhiza*. *Leucheria daucifolia* and *L. pteropogon* are from the arid altiplano some 4000 km to the north. These data *alone* suggest something wrong with the Jara2017 "*L. suaveolens*" sequence. The Lavandero2024 *L. hahnii* sample is identified in GenBank as *L. suaveolens* (per Katinas2022). Its sequences also associate with the southerly species, and its cpDNA sequences are identical to those of Jara2017's *L. hahnii*. The question remains as to what the Jara2017 "*L. suaveolens*" sequences correspond. Specimen misidentification can be ruled out, because the specimen evidently is from the Falkland Islands. Most likely, the sequences represent DNA contamination. Thus, the Jara2017 *L. suaveolens* DNA preparation contained no *L. suaveolens* DNA. The PCR products were weak amplifications of cross-contaminated DNA that yielded the low quality contaminant sequences. The sequences still appear to be distinct from all other species, but, given the evidently low quality of the DNA, these differences may owe to sequencing error.

**viii. *L. integrifolia*.** The cpDNA sequences of the Jara2017 and Lavandero2024 *L. integrifolia* (Phil.) Crisci samples are divergent (Fig. 6), but the dissimilarity is not so great as to rule out polymorphism or sequence error or both. Because of probable sequence errors, I did not include the Jara2017 *L. integrifolia* ITS sequence in the analysis.

**ix. Other species.** Pairwise comparison (against the alignment background) of Jara2017 and Lavandero2024 data for the remaining nominally conspecific sequences revealed that the differences owe mainly to poor sequence quality, especially of the Jara2017 sequences, and especially for ITS sequences. Consequently, several defective sequences were discarded from the present data, because these would have introduced error into the phylogenetic analysis. The remaining sequence pairs mostly do not reveal problems as far as identification is concerned. In fact, the Jara2017/Lavandero2024 cpDNA sequence pairs for six species are essentially identical, ignoring uninformative indel artifacts. Divergences among the other pairs mostly are not greater than what might be expected given the combination of normal infraspecific polymorphism and sequencing error. Jara2017 sequences "of concern" are the ITS for *L. bridgesii* and cpDNA sequences for *L. hieracioides* and *L. integrifolia* (see above). The Jara2017 *L. papillosa* Cabrera cpDNA sequences also seem excessively divergent, but there are no Lavandero2024 sequences for comparison.

### 3. Incongruence between ITS and cpDNA data

Lavandero2024 emphasized the existence of incongruence between their ITS and cpDNA data sets, but this involved incongruent relations at deeper nodes, e.g., relations of the *Spinoliva* and Mutisieae sequences. These same incongruencies were reported earlier by Hershkovitz (2024c). Lavandero2024 did not discuss incongruencies of gene trees among *Leucheria* crown node taxa, though they referred indirectly to the statistically unsupported differences in the relations of the *L.* sect. *Macrobotrys* (viz. *L. floribunda*) sequences, also reported earlier by Hershkovitz (2024b).

While the MP and/or MP bootstrap consensuses and/or ML phylograms show incongruent relations for the ITS and cpDNA data, most of the these involve relations not well supported by the MP bootstrap. Incongruence of the underlying data appears to be minimal in terms of strong bootstrap support for conflicting clades. But there is one clear example of incongruence, which involves the relations of *L. cantillanensis* and *L. salinae*. In Lavandero2024, this incongruence is not evident except very subtly, because they did not sample both *L.* sect. *Lasiorrhiza* subclades. In particular, they did not sample *L.*

*daucifolia* or *L. pteropogon* and did not use the Jara2017 samples of these species or the Jara2017 "*L. suaveolens*" sample. But careful inspection of the Lavandero2024 cpDNA tree does reveal sister relations of *L. cantillanensis* different from those of their ITS tree. The authors evidently failed to notice this.

## 4. Phylogenetic results

The present analysis added to the Hershkovitz (2024b) data some highly similar to identical sequences, as well as the *L. peteroana* sequences. Not surprisingly, therefore, the results do not significantly alter or even add to the major phylogenetic conclusions of Hershkovitz (2024b; here itemized in the Introduction) and Lavandero2024. But comparison of the present results with the previous analyses yields some observations.

The only intersectional relation resolved with strong bootstrap support by at least one data set is the sister-relation between *L.* sections *Polyachyrus* and *Cassiopea*, strongly supported (only) by the cpDNA data. This relation emerged in all cpDNA analyses (Hershkovitz, 2024a–c; Lavandero2024). Lavandero2024 intimated 64% BP support for this relation from ITS alone, but later clarified that this support is for a clade that also includes *L.* sect. *Macrobotrys*. But the sister relation of *L.* sections *Polyachyrus* and *Cassiopea* also persisted in combined data analyses of Hershkovitz (2024a) and, with even more support, in Lavandero2024. This relation did not emerge in the Hershkovitz (2024c) combined data analysis, but this analysis used a reduced sampling of *Leucheria* species. Overall, this is some indication that the ITS data are not strongly incompatible with the cpDNA data on this point.

The Lavandero2024 combined data analysis showed strong BP (92%) support for a sister relation between *L.* sect. *Macrobotrys* (viz. *L. floribunda*) and a clade comprising (as sisters) *L.* sections *Cassiopea* and *Polyachyrus*. Meanwhile, it might be deduced from Lavandero2020's combined data analysis, which excluded *L.* sect. *Polyachyrus*, that they found either 95% or 97% support for a sister relation between *L.* sections. *Macrobotrys* and *Cassiopea*.[41] I did not perform a combined data analysis here, but this relation is consistent with a less well-supported branch in the cpDNA analyses here and in Hershkovitz (2024a, b) and, in fact, the combined data analysis of Hershkovitz (2024a). But the relation is not evident in any of the analyses of ITS alone. Strangely, the relation is evident in Lavandero2024's analysis of ITS alone, but not cpDNA alone, viz. the *contrary* of all of my results. The confounding factor seems to be *Oxyphyllum*, which emerges "all over the place" in the various analyses. Its attraction to *L.* sect. *Macrobotrys* in the ITS MP bootstrap seems to be spurious (see above). But this pair also emerges as sisters, remote from *C.* sect. *Polyachyrus*, in the Lavandero2024 cpDNA ML tree, though with extremely low bootstrap support. However, proper comparison of the present results with those of Lavandero2020 and Lavandero2024 is constrained because they did not make available their alignments and, besides, used different tree construction methods.

The present results "corroborate" the historical ambiguity of evidence regarding the relations of the caulescent taxa to *L.* sect. *Lasiorrhiza*. As noted, monophyly of the caulescent traditional *Leucheria* taxa

---

[41] Lavandero2020 reported that they reanalyzed all of Jara2017 sequences, but they detailed the results only for *L.* sect. *Lasiorrhiza* species. Relations of the remaining taxa were illustrated as strongly supported caulescent sisters "Subclade I" and "Subclade II," one of which *presumably* includes *L.* sect. *Macrobotrys* (viz. *L. floribunda*). The contemporaneously published Pérez2020 designated these same clades as *Leucheria* "Clade II" and "Clade III," with *L.* sect. *Macrobotrys* included in the *latter*. However, these designations are ± the *reverse* of Jara2017, who recognized *L.* sect. *Macrobotrys* as "Subclade I" sister to *L.* sect. *Cassiopea*, which they designated "Subclade II." Thus, Subclades (I + II) of Jara2017 are Clade III of Pérez2020, while "Subclade III" of Jara2017 is "Clade II" of Pérez2020. Subclades I and II of Lavandero2020 must correspond to Clades II and III of Pérez2020, but it is not indicated which is which. *This* is why I published the sectional taxonomy (Hershkovitz, 2024e).

(viz. *L.* sections *Cassiopea*, *Leucheria*, and *Macrobotrys*) was somewhat (Jara2017) to strongly (Lavandero2020) supported in (otherwise flawed) analyses that excluded *L.* sect. *Polyachyrus* and *Oxyphyllum* sequences. Ignoring the relations of the last two, the three traditional *Leucheria* caulescent sections partitioned from *L.* sect. *Lasiorrhiza* in the present ITS MP strict consensus, but with (insignificant) 41/52% BP support (Fig. 2), and also in the ML tree (Fig. 4). But these relations are completely unresolved in the cpDNA analyses. With lesser sampling, the Hershkovitz' (2024a) ITS ML consensus also partitioned the caulescent sections, but the MP strict/bootstrap consensus trees did not. However, one of the cpDNA MP analyses found *polyphyly* of the caulescent section genotypes with 72% support. None of Hershkovitz' (2024b) strict/bootstrap consensus trees resolved these relations. But with even less sampling, a LogDet distance (but neither MP nor ML) analysis partitioned the caulescent sections with 75% BP support (Hershkovitz, 2024c). Lavandero2024's combined data analysis shows 52% BP support for *polyphyly* of the caulescent section genotypes. Again, this is contrary to Lavandero2020, which they failed to mention. Polyphyly is less supported (37%) in their ITS analysis and the relations also are unresolved in the cpDNA analysis. In summary, "monophyly" of the caulescent sections might be insignificantly supported *or* refuted with either ITS or cpDNA, depending on optimization criterion and/or taxon sampling.

At a much lower taxonomic level, current data seem to suggest that the "*L. cerberoana* J.Rémy complex s. str." (viz. this Type plus *L. menana* J.Rémy and *L. cumingii* Hook. & Arn.; see above) and *L. polyclados* (J.Rémy) Reiche are phylogenetically partitioned from remaining *L.* sect. *Cassiopea*. This partition appears with insignificant BP support in both the ITS and cpDNA MP consensus trees (Figures 4, 6). The compatibility of the data sets suggests significant support in a combined analysis, which is what Lavandero2024 shows. The Lavandero2024 combined data analysis also shows their *L. tomentosa* (s. str.) sample diverging next from remaining *L.* sect. *Cassiopea*. This evidently is consequent to the cpDNA and not ITS data.[42] The present cpDNA analysis does show the *L. tomentosa* partition, but it includes also the Jara2017 *L. multiflora* Phil. and *L. senecioides* Hook. & Arn. samples. Remaining *L.* sect. *Cassiopea* include a mix of annual and perennial species whose cpDNA sequences form a strongly supported clade with a relatively long stem branch (Figures 5, 6). But this relation (and associated long stem branch) is not evident in the ITS trees (Figures 2–4). Again, this is peculiar given that ITS is much more variable than the cpDNA loci.

Besides the above, the present results add practically no phylogenetic resolution of interspecific relations within the *Leucheria* sections beyond that reported in Hershkovitz (2024b). The combined data analysis of Lavandero2024 shows some (six by my count) relations within sections supported by high [pseudo-]Bayesian PPs but much lower ML BP support (63–82%).[43,44] This underscores the nature of

---

[42] As noted, owing to sequencing artifacts, I excluded the Lavandero2024 *L. tomentosa* sample from the present ITS analysis and the Jara2017 *L. tomentosa* sample from the cpDNA analysis.

[43] A more *spectacular* example is found in the supplemental tree file of Böhnert et al. (2019; but cf. Böhnert et al., 2022). This analysis focused on "origin and diversification" of the genus *Cristaria* Cav. (Malvaceae) based on three cpDNA loci. Their tree with 29 *Cristaria* terminals was supported by [pseudo-]Bayesian PPs of 0.97–1.0 at *all* 28 nodes. Yet, ML BP support was exceptionally poor: ignoring the stem node (100% BP), only 5/27 nodes have 80–99% support. Three of these five nodes separate a clade of three samples (two species) from the remaining 24 nodes (26 species/varieties). Considering only the latter nodes, 2/24 have 80 and 99% BP, 6/24 have 63–69%, and *16/24 have < 50%*. Thus, the tree is completely ± maximally resolved per PP but very poorly resolved per BP.

[44] Strangely, although *Cristaria* diversification was the focus of this study (with morphology of six species illustrated), the authors referred to lack of interspecific phylogenetic resolution nowhere in their *results*. The authors only alluded to the lack of resolution indirectly in the penultimate sentence of their (very brief) Discussion section ("*The lack of…resolution in interspecific relationships…would be consistent with this scenario*;" Böhnert et al, 2019: 6). The lack of interspecific resolution hardly is surprising given that the analysis used only loci known to evolve too slowly for interspecific resolution. The work really is characterized better as a Malvaceae tribal- to intergeneric-level analysis. Also strangely, the authors illustrated a *fully* resolved interspecific *Cristaria dated*

[pseudo-]Bayesian analysis as an ML optimization algorithm, such that the PPs are not an indication of underlying data support (as for BP support), but rather a biased (viz., systematically overconfident) probability (or confidence interval?) that the branch occurs in the ML tree given the substitution model (Hershkovitz, 2021a).[45,46]

## Discussion

The present work provides an important update to molecular phylogenetics of the *Leucheria* crown group. *Leucheria* sampling in Lavandero2024 was limited. Sampling in Hershkovitz 2024b was greater, but relied on faulty sequences and specimen identifications. This resulted in two separate *Leucheria* phylogenies, both unsatisfactory, with no means of evaluating/reconciling their differences. Since I already had studied Jara2017 in detail (Hershkovitz, 2024a–c), and Lavandero2024 evidently (or supposedly) had not, I was in a unique position to stitch together the two works.

Combining the salvageable sequence data of Jara2017 with the new sequences of Lavandero2024, the present work filled in phylogenetic sampling gaps of the latter, especially for *L.* sections *Cassiopea* and *Lasiorrhiza*. While phylogenetic resolution was not substantially improved, this was neither unexpected, nor the point of the exercise, since the present work is now the *fifth* analysis of sequence data from the same loci in the same genus. But this redundancy at least establishes which nodes in the phylogeny are irresolvable, at least with conventional molecular phylogenetic approaches, but possibly at all. These include relations at deeper levels, viz. intersectional relations, and interspecific relations within sections.

The present analysis helped clarify the identity of several Jara2017 samples, but also demonstrates that most of the Jara2017 identifications are either correct or, if incorrect, that these "species barcode" loci cannot discriminate between taxonomic species. Additionally, this work helped identify errors in both Jara2017 and Lavandero2024 sequences. And, unlike Lavandero2020 and Lavandero2024, the alignments are made available, so that the analyses can be repeated and further refined. Meanwhile, Lavandero2024 advertised their intention to investigate *Leucheria* phylogeny using certain phylogenomic techniques. This approach might render moot some of the finer points of locus-specific analysis discussed here. But

---

phylogeny – with a topology inexplicably very *different* from that in their supplemental file. They even indicated the calculated divergence dates for *Cristaria* nodes unsupported by their data (Böhnert et al., 2019: Fig. 2D). As far as I can determine, the only way the reader could detect the arbitrariness of the Fig. 2D interspecific chronogram would be to study the supplemental treefile. Böhnert et al. (2022) later published a phylogenomic analysis whose BPs *strongly* conflicted with *five of six* of the 80+% BPs in the Böhnert et al. (2019) cpDNA analysis, which Böhnert et al. (2022) failed to mention.

[45] Hershkovitz (2021a) alluded to an empirical concordance between BPs > 70% and PPs ≥ 95%. Hershkovitz (2024b) likewise considered that PPs ≥ 95% were "significant" in some sense. But in view of subsequent discussion in Hershkovitz (2021a) and its empirical "corroboration" in Böhnert2019 and the case of PPs discordant in Lavandero2020 and Lavandero2025 (discussed above), no general statistical relation between BPs and PPs can be supposed, and no PP value can be afforded an inherent credibility value in the absence of corroborating evidence (contra Jara2017, Swenson et al., 2025, among many others).

[46] This point and, in fact, phylogenetic theory and computational methods generally, evidently are poorly understood in the plant systematics community globally. Jara2017 and Swenson et al. (2025), certainly among other multi-authored works published in peer-reviewed international systematics journals, analyzed DNA sequence data using *only* [pseudo-]Bayesian estimation. With no theoretical justification (or understanding), Jara2017 accepted 90% PP as credible, and Swenson et al. (2025) 85%. Theoretical and empirical results elsewhere (Böhnert2019) render these criteria as purely arbitrary. Hershkovitz (2021a) attributed the popularity of [pseudo-]Bayesian estimation to the bandwagon effect and to the ready availability of software that rapidly yields "support" levels that are both aesthetically satisfying and higher than BP support levels. The method is essentially an "opiate of the masses."

phylogenomics presents different theoretical challenges, a few of which will be discussed here in relation to the study of *Leucheria*. A theme emerging repeatedly in this discussion of recent *Leucheria* research is that, despite what current empirical systematic and evolutionary literature might superficially suggest, solutions to scientific questions cannot be "manufactured" by protocols and software. Solutions are validated only by scientific theory that, unfortunately, protocols and software…and hence the literature …tends to mask or even conceal. In a theoretical vacuum, protocols and software are as likely to misinform as inform. At the same time, with adequate theoretical understanding, they can serve their intended purpose as powerful and efficient discovery tools.

## 1. Molecular evidence for species taxonomy in *Leucheria*

Crisci (1976) and Katinas2022 presented rather different species taxonomies of the genus, the latter tending to lump species recognized as distinct in the former. As Lavandero2024 noted, however, both taxonomies were based on study of herbarium specimens only, and no more than a few of those actually were collected by those authors. The present combination of Jara2017 and Lavandero2024 data offers possible insights on the species taxonomy of *Leucheria*. These data must be viewed cautiously in view of questions of specimen identification (Lavadero2024), possible sequencing artifacts, and, most of all, the *theoretically predicted* and empirically established lack of 1:1 correspondence between DNA sequences and species taxonomy (Hershkovitz, 2021a; see also Nicola et al., 2019; Böhnert et al., 2022; or, for that matter, practically any recent study with multiple interspecific samples; see also later discussion). Nevertheless, in this case, combining the available sequence data mitigated marginally the principal challenge to molecular-level understanding of interspecific evolution, viz. inadequate sampling, and the results help orient further research.

As noted, Apodaca et al. (2021) and Katinas2022 lumped into *L. tomentosa* a total of ten species recognized by Crisci (1976) and 24 total species Types (apospecies sensu Hershkovitz, 2019a, 2021a, 2022). These are all of the *Leucheria* species described as annuals by Crisci (1976) and other authors, including Lavandero2024. Following Katinas & Forte (2020), Katinas2022 described them as perennials. In the Introduction, I reported observations of decidedly semelparous annual behavior. In any case, as established by Jara2017, Hershkovitz (2024b), and Lavandero2024, genotypes of the taxa so lumped are polyphyletic, some belonging to *L.* sect. *Cassiopea* and others to *L.* sect. *Leucheria*. Thus, despite their morphological similarity/intergradation, their synspecificity (sensu Hershkovitz, 2019a, 2021a, 2022) is questionable.

The current data show that specimens of annuals identified by Lavandero2024 as *L. tomentosa*, *L. cerberoana*, *L. oligocephala*, and *L. tenuis* pertain to *L. sect. Cassiopea*, along with specimens identified by Jara2017 as *L. cumingii, L. menana*, *L. multiflora* and *L. senecioides*. The specimens that Jara2017 identified as *L. oligocephala* and *L. tenuis*, whose DNA maps them to *L.* sect. *Leucheria*, presumably were misidentified (cf. Lavandero2024; see above). The sequences of the *L.* sect. *Cassiopea* samples are unique in either or both of the data sets, though I cannot verify whether the differences represent true polymorphisms or sequencing artifacts.

Within *L.* sect. *Cassiopea*. the sequences of *L. cumingii*, *L. cerberoana*, and *L. menana* form a clade in both the ITS and cpDNA analyses, and sequences of the last two are identical or hardly different. The Types of all three are from Coquimbo. This is some evidence that these, or at least the last two, are synspecific. The data seem to suggest that the remaining annual species in this section pertain to one or more distinct species. But the evidence is nuanced. Both Jara2017 and Lavandero2024 showed/reported that genotypes of the annual species of *L.* sect. *Cassiopea* were polyphyletic. Logically, this would render

them heterospecific. But both analyses, especially the former, are problematic in this respect.[47] In the present analysis, only one branch in the cpDNA tree (and none in the ITS tree) with > 70% BP supports polyphyly (Fig. 5). In this case, the sequence similarity (Fig. 6) of the annuals *L. oligocephala* and *L. tenuis* with certain *L.* sect. *Cassiopea* perennials and their sequence divergence from other *L.* sect. *Cassiopea* annuals raise the specter of earlier interspecific gene flow and introgression (and/or lineage sorting). This complicates taxonomic diagnosis. It also means that indeed the obligately annual phenotype/behavior[48] could have a single *origin* in *L.* sect. *Cassiopea*, even if their "species barcode" genotypes are polyphyletic (see later discussion). Meanwhile, the sequence evidence suggests that annual plant samples pertaining *L.* sect. *Leucheria* and identified by Jara2017 and Lavandero2024 as pertaining to four different species pertain to just two, which appear in both data set analyses as sister: *L. glandulosa* and *L. glabriuscula*.

In summary, on the surface, the data appear to support multiple origins of the annual habit. But among the alternatives scenarios is one in which the annual forms are indeed descendents of two annual ancestors, and that apparent polyphyly of the *L.* sect. *Cassiopea* annuals is consequent to introgression and/or lineage sorting. In fact, ancient hybridization cannot be ruled out to explain the sectional level polyphyly of the annual habit, either, so that the annual habit evolved only *once* in *Leucheria*. Phylogenomic approaches and additional sampling could test these alternatives, in particular whether other loci otherwise share common or disparate histories.

Besides the above, the sequence data offer little in the way of taxonomic clarification. Genotypes within the subclades of *L.* sections *Leucheria* and *Lasiorrhiza* sort differently in the two data sets (Figures 4, 6), but this may reflect a "normal" pattern of polymorphism found among very closely related species or, alternatively, sequencing artifacts or, perhaps most likely, both. For example, Katinas2022 included *L. garciana* J.Rémy in *L. gilliesii* Hook. & Arn.. The ITS data are consistent with this opinion, but the cpDNA data are not. For ITS, the *L. lithospermifolia* Reiche sample has the same sequence as the *L. rosea* Less. sample. For cpDNA, it has the same sequence as the *L. garciana* samples. Given the low divergence, I draw no conclusion from this. Polyphyly of the two *L. hieracioides* cpDNA sequences might be explained different ways, but speculation is pointless in the absence of additional data.

Within *L.* sect. *Lasiorrhiza,* in both data sets, the Jara2017 *L. pteropogon* sample clusters in the same subclade as the *L. daucifolia* sample, but the genotypes are distinct. The Jara2017 "*L. suaveolens*" sample clusters here also, but I ignore this here (see above). But Katinas2022 merged *L. pteropogon* into *L. salinae*. Moreover, Katinas2022 classified *both* the Jara2017 *L. daucifolia* and *L. pteropogon* specimens as *L. salinae*. Katinas2022's classification renders a distribution of *L. salinae* from ca. 18–36S, and from the western slope of the Andes eastward to north central Argentina. But when segregated, the distributions of these species appear distinct (Crisci, 1976). *Leucheria pteropogon* sensu Crisci (1976) is

---

[47] I elaborated elsewhere (Hershkovitz, 2024b) the problems of Jara2017, besides specimen identification, discussed here. In the case of Lavandero2024, each of their two separate and single combined data trees supports polyphyly of the *L.* sect. *Cassiopea* annual species genotypes, separated by two branches with > 70% BP support. At least this is the way it appears; the graphic resolution, especially of the cpDNA topology, is so compressed as to be undecipherable. In any case, the *three* topologies appear to be mutually incompatible. While it is theoretically possible for a well-supported branch in a combined data topology to conflict with *both* well-supported conflicting branches in separate data topologies, it is not common operationally. The case of Jara2017 owed to "noise" in the separate data analyses that partially canceled out in the combined data analysis, yielding a novel topology. Since Lavandero2024 did not make their alignment available, I cannot identify the cause of the mutual incompatibility of their three topologies.

[48] Physically defined phenotypes are consequent to particular ontogenetic *behaviors*. Maturana & Mpodozis (2000; Mpodozis, 2022) used the term "ontogenetic phenotype" to capture this form/function duality, which is universal (e.g., the relationship between mass and energy). Just as importantly, behavior is equivalent to *evolution*, viz. an irreversible change in course. See also later discussion.

distributed from the Chilean altiplano eastward to the highlands of north-central Argentina, whereas *L. salinae* sensu Crisci was restricted to the Andes of central Chile and Argentina. Meanwhile, *L. daucifolia* is distributed in the highlands of Bolivia and southern Peru, except for the northernmost Chilean collection that Katinas2022 classified as *L. salinae* (see below). Notably, Katinas2022's classification created multiple latitudinal disjunctions in the documented distribution of so-circumscribed *L. salina*: in between ca. 18S and 22S (ca. 450 km), ca. 22S and 28.5S (ca. 700 km), and ca. 28.5S and 31S (ca. 250 km). Thereafter, the distribution is continuous between ca. 31–36S.

The relevant molecular evidence is complicated, but it does not support Katinas2022's opinion. The ITS data show Jara2017 and Lavandero2024 genotypes identified as *L. salinae* as polyphyletic (Figures 2, 4), though within the same *L.* sect. *Lasiorrhiza* subclade. One sample is sister to *L. daucifolia*, with the *L. pteropogon* sample sister to these. The other sample is sister to *L. cantillanensis*. These data alone do not strongly refute Katinas2022's opinion, given evidence for similar infraspecific ITS behavior in other taxa. But in the cpDNA tree, both *L. salinae* samples, along with *L. cantillanensis*, map to the other *L.* sect. *Lasiorrhiza* subclade. The *L. daucifolia* and *L. pteropogon* samples remain as sisters in the original subclade.

From Jara2017, the provenances of the *L. daucifolia* and *L. pteropogon* samples (classified as *L. salinae* by Katinas2022) can be deduced. The former is *Arroyo 84-845* (CONC) from Chile's Arica & Parinacota Regions at ca. 18S, near the Bolivian border. Multiple specimens accepted by Katinas2022 as *L. pteropogon* map to Bolivia at the same ‚latitude, but some 50–100 km to the east. The latter is *Arroyo 97-016* or *97-025* (CONC) from Chile's Antofagasta Region near the Argentinean border at ca. 22S, ca. 250 km NW of the *L. pteropogon* Type locality in Argentina's Salta Province. Setting aside the question of *L. salinae* taxonomy, no other *Leucheria* species occur in these respective regions. Katinas2022 classified all Bolivian and Peruvian collections as *L. daucifolia*, and all altiplano Chilean and Argentinean collections as *L. salinae*. Meanwhile, the Jara2017 *L. salinae* sample is from central Chile at ca. 36S. I suspect that the Lavandero2024 sample also is from central Chile. The *L. salinae* Type is from ca. 32S.

The molecular data suggest that the central Chilean samples of *L. salinae* pertain to an ancient hybridization event between an altiplano lineage that gave rise to the *L. daucifolia* and *L. pteropogon* types, and a plant pertinent to the other *L.* sect. *Lasiorrhiza* subclade, which comprises the Patagonian species. This tryst could have been consequential to dispersal or to latitudinal climate-mitigated migrations. This hybrid later diverged as *L. salinae* and *L. cantillanensis*.[49] Such a hybridization event would explain the morphological similarities between *L. salinae* and the more southerly altiplano *L. pteropogon*, leading to Katinas2022' taxonomic submersion of the latter. Notably, while Katinas2022 included *L. pteropogon* in *L. salina*, at the same time they distinguished *L. salinae* from *L. daucifolia*. The molecular data suggest that *L. pteropogon* should be distinguished also from *L. salinae*. This is the most parsimonious explanation given the data. Other scenarios are possible, e.g., involving lineage sorting of alleles from a genetically polymorphic ancestor. It would be useful to genotype more samples from throughout the respective ranges and, especially, from Type localities.

Lavandero2024 concluded that the molecular evidence refutes Katinas2022's argument for merging of *L. millefolium* Dusén & Skottsb. into *L. purpurea* (Vahl.) Hook. & Arn. The ITS data indeed separates these by a strong BP (Fig. 2). Meanwhile, the ITS sequences of *L. purpurea* are identical to those of the *L.*

[49] *Leucheria cantillanensis* is endemic to a locality in central Chile's coastal ranges at ca. 34S. It is the only species of *L.* sect. *Lasiorrhiza* that occurs in the coastal ranges (Lavandero2020). Lavandero2020 suggested, effectively, that it arrived at this locality when alpine Andean taxa descended in elevation and spread across central Chilean valleys during Quaternary glacial cycles. It later would have become stranded atop its coastal range perch, where it differentiated. Alternatively, it seems that it could have arrived easily earlier via wind or bird dissemination of its lightweight plumose achenes. But there is no evidence available to support or refute either hypothesis.

*eriocephala* Speg. samples. But the cpDNA sequences of samples of all three species are identical (Fig. 6). The ITS data are intriguing, because *L. millefolium* and *L. eriocephala* are sympatric and have similar distributions, with *L. purpurea* parapatric to the south of both (Katinas2022). Katinas2022 evidently regards *L. millefolium* and *L. purpurea* to be the most similar pair on morphological bases, but this is not clear to me from the descriptions there or in Crisci (1976). It should be noted, however, that the ITS divergence in this case is much less than that between the intersectional annual species divergence and in the *L. salinae* incongruence case. Also, there is only one ITS sample for *L. millefolium*. Given the idiosyncrasies of evolution at this low level of molecular, morphological, and geographic divergence, I do not consider the meager ITS evidence to be definitive on this taxonomic matter. If the evidence is considered sufficient to justify separation of *L. purpurea* and *L. millefolium*, would it not, in turn, justify *merging* of *L. purpurea* and *L. eriocephala*?

While the available genetic evidence for *L.* sect. *Lasiorrhiza* species is inadequate for resolving phylogenetic or even taxonomic questions, it actually is positive evidence that the species themselves have not *individuated* as historically independent lineages. This hardly is unprecedented or theoretically unexpected. Among analogous taxa are *Oriastrum* Poepp. subg. *Egania* (J.Rémy) A.M.R.Davies (Mutiseae; Hershkovitz, 2021a) and *Nassauvia* Comm. ex Juss. subg. *Strongyloma* (DC) Cabrera (Nassauvieae; Nicola et al., 2019). Both taxa comprise genetically closely related Andean acaulescent arctic/alpine herbaceous perennial Mutisioideae lineages that are highly diverged genetically from their nearest relatives.[50]

The case of *N.* subg. *Strongyloma* is especially relevant, because the five accepted morphologically-diagnosed species share the same southern Patagonian geographic range as several accepted species of *L.* sect. *Lasiorrhiza*. Moreover, species ranges in both taxa overlap, viz., there is considerable sympatry. Nicola et al. (2019) sequenced ITS and two cpDNA loci for 44 individuals of the five species and found that numerous individuals were polymorphic for plastid haplotype or, more often, nucleotype, with up to nine nucleotypes within an individual. But they also found that the haplotype and nucleotype distributions and trees did not remotely coincide with the species taxonomy or, for that matter, with each other. They reasoned that this was consequent to the "usual suspects," viz. hybridization and/or lineage sorting, in conjunction with historical ecological factors (e.g., glaciation cycles, anemophily). Morphological homoplasy and plasticity might be added to this equation. The present results suggest that a similar approach to the study of *L.* sect. *Lasiorrhiza* species would be warranted, especially given the evidence for past hybridization and/or lineage sorting between the Patagonian and northern ChFR species.

But the data underscore the *significance* of the aphorism that "gene trees are not species trees," not because of their incongruence, but because genomes and morphological species are ontologically and behaviorally distinct entities sometimes only loosely adherent to their material milieu. It also is critical to note that evolutionary studies of species complexes such as that of *L.* sect. *Lasiorrhiza* transcend the conventional boundary between taxonomic species, however fuzzy, and (meta)populations of individuals whose a priori taxonomic identity likewise may not be clear. Available statistical approaches to such studies mix and match approaches from phylogenetics and population genetics (e.g., Nicola et al., 2019), each laden with assumptions that may not be biologically realistic generally or in specific cases. And the most fundamental, yet biologically most unrealistic assumption of the methods is that biological evolution is a *stochastic* process (Hershkovitz, 2021a).

[50] From a genetic divergence standpoint, there is no shortage of similar examples of lineages of other angiosperms, other growth forms, and/or habitats.

## 2. Data set incongruence

As indicated in the above discussion, the present results yielded examples of incongruence between the ITS and cpDNA trees of *Leucheria*. Hershkovitz (2024a–c) discussed some of these and others among related Nassauvieae. Lavandero2024 later discussed data set incongruence as a factor confounding phylogenetic analysis of these taxa. While not citing the above studies, they remarked that, among five earlier works that they did cite, "…***not all****…documented the topological incongruencies between the chloroplast and nuclear ribosomal cistron* [**emphasis mine]**." Technically, they were correct. Only four of the five (viz. not "*all*") of the works they cited documented incongruence. The only one that did not was Jara-Arancio et al. (2017b ["2018"]). And another, Jara2017, *documented* incongruence, but they dismissed its significance.[51] Lavandero2024 also did not mention that their *own* earlier *Leucheria* analysis, Lavandero2020, did not document incongruence, since they only performed a combined data reanalysis of the Jara2017 data, whose incongruence they did not acknowledge.

To detect incongruence between their ITS and cpDNA data, Lavandero2024 performed the Incongruence Length Difference Test (ILD) as implemented in PAUP, citing only the latter as the reference. Later, they remarked that "…*only a few*… [previous]…*studies explicitly performed tests to evaluate incongruence between nuclear and chloroplast partitions*." Besides overlooking their own earlier failures to evaluate/detect incongruence, the authors also apparently overlooked a half-century of theoretical/empirical literature on the subject.[52] Most notably, Barker & Lutzoni (2002; 515 Google Scholar citations; cf. Planet, 2005 ["2006"]; Wang et al., 2014) corroborated reports of a high incidence of Type I and, more frequently and severely, Type II ILD test error.[53] They concluded that the ILD is not

---

[51] Hence this incongruence was ignored also in Pérez2020, which based upon Jara2017.

[52] In the context of cladistic phylogenetics, the history traces to Mickevich & Farris (1981). This study and the advent of combined morphology/molecular analysis led Farris et al. (1994, 1995; not cited by Lavandero2024) to develop the ILD test. Since PAUP aspired to be a complete and all-purpose phylogenetic analytical platform, it agnostically incorporated this among several other outsourced routines. But almost immediately, empirical researchers began to report anomalous ILD results.

[53] Research on ILD has focused on its poor performance, operational explanations thereof, and possible operational repairs. But the usually unstated underlying theoretical reason for ILD failure should be intuitively obvious to anyone with a working *conceptual* understanding of phylogenetic theory and math. The ILD bases on the differences between MP tree lengths (X; the independent variable) and their frequency distribution (Y; the dependent variable). But while tree *length* distribution is linear, tree *topology* distribution is not. First, it must be emphasized that phylogenetic trees are not real physical trees, but *metaphorical* trees derived in hypergraph space that approximate the phylogenetic process for purposes of mathematical analysis of the latter. While a single cladogram (or dendrogram) is planar (viz., two-dimensional), this metaphorical tree space is *not*. The *mathematical* space in which cladograms are derived is *hyperdimensional* or, more precisely a *hypercube* that includes all possible (two-dimensional) *Steiner tree* (~cladogram) resolutions. Hypercubicity of cladogram space can be appreciated by considering that all possible direct paths between four terminals (–A, –B, –C, –D; the *minimally* phylogenetically informative "four-taxon" case) define three sides of a cube. Adding another terminal (–E) completes the path cube. Adding more terminals exponentially generates more cubes whose edges connect all terminals. Flattening the hypercubes in every possible way creates the distinct cladograms in which paths between most terminals necessarily are shared by "shortcuts" through shared nodes. Adding character data has two effects. First, it yields a *Buneman tree*, which is the hypercube subset that contains the set of cladograms compatible with each distinct empirical character data pattern. Second, the data can be used to weight the edge lengths, hence also path lengths. This yields *phylograms* (or, alternatively, ultrametric chronograms). Because of unequal edge lengths, phylogram and chronogram space is not hypercubic, but rather *hyperhexahedric*. [The Splitstree program (Huson & Bryant, 2006) permits visualization of three-dimensional *polyhexahedrograms* (3D phylograms) generated by split decompositional analysis of nearest neighbor 4-taxon splits.] However, because phylogenetic cladistic analysis is weighted by the data, even the consequent cladograms effectively are rescaled phylograms. Thus, every two-dimensional cladogram (divergence X time in *phylogenetic* space) occupies a parametric data pattern *dimension* in the *mathematical* space in which cladograms are evaluated and "optimized," such that all possible cladogram

an unbiased measure of phylogenetic congruence and that, operationally, "…[analytical/interpretative] *decisions based on the ILD would be misleading in a large proportion of cases*" (Barker & Lutzoni, 2002: 51).[54] *This* is the reason why ILD fell out of (dogmatic) favor and why more recent studies (like those cited by Lavandero2024) do not "explicitly perform" this test. Thus, while Lavandero2024 seemed to try to manifest a sort of theoretical enlightenment by their invocation of ILD, they demonstrated quite the opposite.[55]

The above discussion, like most of the relevant literature itself, addresses the issue of *topological* incongruence only. The present results also highlighted marked differences between the data sets in "branch lengths," or the amount of observed or estimated phylogenetic change between inferred ancestors and observed descendents. This, too, can affect ILD and bootstrap results because of data swamping. In particular, it may not be clear the degree to which smaller but possibly "significant" support for a shorter branch in one data set is incompatible with much stronger support for an incompatible longer branch in another. Branch length incongruency (and its underlying cause) is a case where the ILD will yield a Type I error. Scrutiny of ΔBP is a crude way of detecting it.

But branch length incongruency between data sets is more problematic in downstream analyses involving transformation of phylograms into ultrametric trees, viz., statistical branch length correction for

---

*lengths*, while *distributed* linearly, are not *related* linearly for comparative purposes. Cladogram similarity, as opposed to length similarity, is a function of the number of shared tree edges (internal branches), or the number of *path changes* (rearrangements) required to convert one tree into another (cf. Goloboff, 2008). The cost of moving a branch is not simply its branch length subtracted from one point and added to another, because the move creates "shortcuts" through other Buneman hyperhexahedric tree dimensions. Synapomorphies in one resolution become homoplasies in another. Thus, cladograms with identical or (for ILD purposes) statistically similar (or different) lengths may be very different (or similar) in their hyperhexahedric Buneman space. *This* is why the ILD is prone to Type I and Type II error. Another way to explain this problem is to recall that an MP tree is an ML tree that assumes that each character evolves independently (Swofford et al. 1996). Like MP tree lengths, likelihoods for different ML topologies based on the same data and model are distributed linearly. But they *cannot* be compared statistically, because the topology itself is a parameter incorporated into the likelihood calculation (Yang et al., 1995). Thus, the models are not nested for purposes of likelihood ratio tests (except in the trivial zero-versus-positive branch length case), and the statistical significance of the likelihood difference among topologies cannot be determined. The same logic thus applies to MP trees conceived as ML trees: the topology itself is a parameter of the length, so the statistical significance of ILDs cannot be determined. The above two explanations are identical if it is understood that *each* possible fully resolved topology defines a nonlinear parametric *dimension* that cannot be nested within any other fully resolved topological dimension  Meanwhile, models of *linear* parametric distributions can be nested and reductively combined (e.g., as in ModelTest).

[54] In fact, *single* data set partitions are known to "fail" the ILD test, which I personally have confirmed in the course of my research.

[55] With these caveats in mind, the ILD can be and often is applied *heuristically*, but it generally does not substitute for subjective comparison of separate versus combined data trees (e.g., Wang et al., 2014). Hershkovitz (2021a) performed the ILD test, which reported incongruence. He then undertook a tedious branch-by-branch analysis of separate data trees and recorded ΔBP between the separate and combined analyses. In this particular data set, the *number* of branches with ≥70% BP increased overall by about one third in the combined versus separate analyses. This is indicative of data set congruence/complementarity at *these* nodes, *overall* significant ILD incongruence notwithstanding. At the same time, support for *some* branches in one or the other separate data set *decreased* when data were combined, indicating incongruence. But there is no *formal* statistical procedure that, in the context of large trees, tests *branchwise* data set incongruence. Also, as Hershkovitz (2021a) noted, the "significance" of the BP is *relative,* because it depends upon ($BP_{max}$ - $BP_{max'}$), where $BP_{max'}$ is the maximum BP of any *incompatible* partition in any alternative topology. ($BP_{max}$ - $BP_{max'}$) ranges in between ~zero and ~50%, such that the difference between, e.g., $BP_{max}$ 51% and $BP_{max'}$ 49% is "insignificant," whereas that between 51% and 1% is large enough to render credible the 51% partition. However, the principal conclusion of Hershkovitz (2021a) was that the value of statistical analyses/tests in phylogenetics is purely *heuristic*, because statistics presume a strictly *stochastic* evolutionary process, whereas Hershkovitz (2021a) argued that it is an *idiosyncratic* process.

purposes of macroevolutionary ecological and/or geological age analyses. Let us assume that the topologies themselves are completely congruent. When data are combined, the branch length correction effectively sums/averages the incongruent branch lengths according to their data weight. This will yield "corrected" branch lengths that are (probably statistically significantly) different from the similarly "corrected" branch lengths for one or both of the original data sets. Is this a legitimate operation? In statistical terms, the question can be phrased as whether or not the original incongruent branch lengths are drawn from the same linear distribution of branch lengths generated by some Markov process. Alternatively, the branch length distributions themselves may differ between data sets and even among branches generated by a single data set. However, I generally reject the notion that evolution is stochastic, even though an assumption of stochasticity may have heuristic value. The point here is that, in empirical practice, ultrametric trees commonly are generated using combined data sets without reference to manifested branch length incongruence. I will not elaborate further on this subject here, other than to point out that branch length incongruence among/within data sets also is a parameter of incongruence, and it seems to be generally unappreciated/ignored.

## 3. Phylogenomics a propos *Leucheria* systematic and evolutionary study

Lavandero2024 advertised their presumably ongoing phylogenomic research on *Leucheria* using the targeted nuclear gene approach. This is one of the four currently popular genomic mining approaches applied in phylogenomics, and the one currently considered most broadly applicable and reliable (Yu et al., 2018).[56] My first thought, however, is to wonder how a research group manifestly deficient in both theoretical and practical aspects of *single*-locus phylogenetic analysis can be proficient in simultaneous analysis of hundreds of loci. The partial answer is "artificial intelligence" that is supposed to obviate certain pitfalls of conventional analysis.

For example, Jara2017 followed established mechanical protocols, but lacked the theoretical knowledge and practical experience to avoid/detect PCR contamination, sequence artifacts, and alignment errors. In modern phylogenomics, all protocol steps in between DNA extraction and alignment are outsourced to machines and analytical programs. Phylogenetic analysis effectively also is outsourced to third-party software. Phylogenomic "research" *technically* is less intellectually demanding than conventional molecular systematics.[57] Yet, perhaps even more so than conventional molecular systematics before it, phylogenomics literature currently is *sensational*, viz. has high sensorial impact, is afforded high a priori credibility, and is socioeconomically and politically disproportionately lucrative for its authors. This, perhaps, is because it imparts a sense of the authors' exceptional scientific knowledge, understanding, abilities, and accomplishments across multiple scientific disciplines, viz. molecular genetics, genomics, evolutionary and phylogenetic theory, hypergraph theory, statistics, as well as classical systematics and ecology.[58] But by virtue of the "magic" of AI and ever less stringent coauthorship qualifications, phylogenomics, even more than molecular systematics research, requires *no*

[56] Obviously, these approaches will be rendered obsolete by fast, cheap, automated whole genome sequencing and assembly. This will all but eliminate the researcher laboratory effort and render at least the molecular genetic aspects of phylogenomics as a purely AI computational paradigm.

[57] This is to say that, as demonstrated in Hershkovitz (2024a), even the most technologically sophisticated conventional molecular systematics still requires a greater degree of researcher interaction. This was even more so before the advent of such conveniences as automated DNA sequencing, modern DNA extraction kits, PCR cloning kits, automated DNA substitution model selection software, etc.

[58] For some publications, one would imagine also expertise in climate science.

scientific abilities or accomplishment. It does not even require that an author actually contributed to the research or understood or even actually read the paper.[59]

Hershkovitz (2021b) summarized a *few* of the (unsolved) theoretical and practical challenges of phylogenomic evolutionary reconstruction. To a degree, the finer-scale details of genome evolution may be red herrings. While phylogenomic reconstruction methods usually erroneously assume that sequences of each locus or, more commonly, all loci together, evolve according to a uniform first-order Markov *base* substitution process, it is possible that violations of this assumption are *collectively* innocuous. In the case of single locus analysis, assumption violations will bias the results along some vector. But in the case of 100 or more loci, these error vectors might converge on a normal distribution, as though the substitutions themselves were indeed IID. Consequently, phylogenomic cladograms and bootstrap values may be statistically valid approximations.

But not so fast. A complete review of the potential pitfalls of simplistic phylogenomic analysis is beyond the scope of the present work. But a few abstract observations are in order. Perhaps the greatest challenge to phylogenomics is the conceptual straightjacket imposed by (once vanguard but now obsolete) metaphorical "tree thinking" as naively conceived and dogmatically marketed in the context of conventional molecular systematics by Baum & Smith (2012; cf. Hershkovitz, 2019b).[60] More precisely, cladogram-thinking constrains the conceptualization of the evolutionary process into just two dimensions. This conceptualization is less problematic for lineages that are irreversibly evolutionarily highly divergent and individuated. In these lineages, *ideally*, the genome historically was completely "contained" within the lineage, such that at least the confounding processes of horizontal gene transfer and lineage sorting might be discounted. Cladogram-thinking also is less problematic for analyses using fewer and better-studied loci, whose evolutionary dynamics across broader phylogenetic space can be documented, and whose phylogenetic incongruencies can be detected relatively easily. Yet, conventional molecular systematic evidence does sometimes demonstrate polyphyly of the collective loci within well isolated lineages. At lower taxonomic levels, one hardly needs phylogenomic or even molecular systematic evidence to appreciate reticulate (viz. noncladistic) evolution consequent to, e.g., hybridization and lineage sorting. This was evident decades before DNA technology came along.

Phylogenomics takes unjustified assumptions to a new level. It effectively idealizes the genome as equivalent to, e.g., a *single* well-studied locus that is *proportionally* orders of magnitude more informative but not at all more "misinformative." This is absurd, as any late 20th Century undergraduate molecular genetic/evolution text should make clear. Perhaps it is unappreciated that the *reason* for the abrupt shift from conventional single-locus to phylogenomic 100-locus+ analyses does not owe strictly to technological advancement. The reason that more loci were not exploited conventionally is that they have not evolved sufficiently (as relatively[61]) *predictably* for both conventional laboratory and phylogenetic analytical purposes. This point must not be lost upon wannabe phylogenomicists. As phylogenomics increases linearly the number of analyzed loci, it increases geometrically both the a priori and a posteriori per-locus ignorance of the evolutionary dynamics. This is because, unlike the case of "standard" loci, the massive size of the data logistically prohibits per-locus scrutiny while, at the same time, provides little per-locus comparative basis for such scrutinization.[62]

---

[59] Science is supposed to proxy for human rationality, but, in this way, the institution of science is converging on *irrationality*.

[60] To wit, "*species to me are clades*" (D. Baum, written comm., 25 Dec. 2015).

[61] It should be clear from this and countless other theoretical and empirical works that "standard" loci do not evolve *completely* predictably and, as noted, according to simplistic substitution models.

[62] The importance of such scrutiny should be rendered clear by this work, Hershkovitz (2024a–c), and countless other theoretical and empirical works.

At the same time, while increasing locus number increases empirical uncertainty, it also exponentially increases the analytical complexity, because each locus defines its *own* hyperhexahedric Buneman *galaxy*. Yet, researchers are more or less blind to the computational process that reduces these hyperhexahedric Buneman galaxies into (asymptotically) a single optimal[63] cladogram. But because cladogram-thinking trains the mind to *idealize* a cladogram, researchers *inherently* are satisfied with this result and therefore disinclined to question it. The overall effect is that phylogenomic approaches preemptively bias against discovery of phenomena that phylogenomic mining renders increasingly likely. Then reviewers, editors, and readers likewise do not scrutinize the data.

The phenomena supposed to confound both phylogenomic and single-locus phylogenetic analysis include, besides homoplasy, *intergenomic* horizontal gene transfer (HGT) and lineage-sorting of ancestral allelic polymorphism, whether consequent to HGT or mutation. Hershkovitz (2021a) suggested another phenomenon: vertical gene transfer (VGT). Actually, VGT comprises *infragenomic* HGT and *nonallelic* lineage sorting. Phylogenomic and conventional molecular phylogenetic analysis assumes strict intergenerational Mendelian allelic inheritance at each locus. But genomic-level recombinatorial processes themselves can alter sequences at a locus and otherwise modify/rearrange alleles. Hershkovitz (2021a) mentioned gene conversion and transposition via mobile elements. Mitochondrial plasmids are another route. Thus, the hyperhexahedric Buneman galaxies themselves interact in time and do not evolve independently.

Setting aside the complexity of metaphorical nonindependently-evolving hyperhexahedric Buneman galaxies, Hershkovitz (2021b) suggested possible bias in phylogenomic studies that deliberately target loci that are functionally related at the phenotypic level, e.g., Hancock et al. (2018) and developmental stage-specific transcriptome mining. Bias may be less in the in the case of arbitrarily selected loci. But even then, bias may occur for several theoretically predictable and empirically demonstrable reasons. This discussion, again, is beyond the scope of the present work, except to note that empirical evidence for bias has emerged in disparate results generated using different data subsets and/or the same data but different reconstructive methods. This is because, under unbiased conditions, different methods should yield the same results. The point is that, in phylogenomic analysis, as in any statistical analyses, the onus is on researchers to demonstrate that their results are unbiased.[64]

Finally, phylogenomics brings to front and center epistemological questions raised only incidentally by conventional molecular systematics. This is the ontological relation between genomes and taxa. I have emphasized here and elsewhere (Hershkovitz, 2019a, b; 2021a) that the ontological distinction between genes and species prohibits ontological equivalence even between congruent gene and species trees or, more generally, organismal-level taxon trees. But even so, the relationship between conventional phylogenetic marker gene divergence and taxon divergence is generally considered to be *correlative* and neither causal, nor even diagnostic.[65] The whole genome is another matter, because the reductionist doctrine of genetic determinacy indeed posits that species differences are "coded by" DNA-

[63] The "optimal" path in *phylogenetics* (as opposed to mathematics/statistics) is the arithmetically "shortest" or philosophically "most parsimonious" reconstruction of evolutionary history given particular a priori *epistemological* and *biological* (as opposed to purely mathematical/statistical) rules. Thus, phylogenetic reconstruction is an epistemological and biological, rather than mathematical/statistical, problem. Computational output of phylogenetic – and, in fact, all evolutionary and ecological results – is meaningless except with reference to the underlying epistemological and biological rules.

[64] I think I just heard a huge global wave of laughter.

[65] But see Wolf et al. (2013), who discussed existing and new evidence for a diagnostic relationship between "biological species," viz. comprising interfertile individuals, and compensatory base substitutions in rDNA ITS2 secondary structure. See also Reid & Carstens (2012) and Qin et al. (2017) and citing references for discussions of species diagnostics based on phylogenetic marker divergence.

level differences at loci dispersed throughout the genome. This still widely-accepted but false doctrine implies that genome and species trees indeed should be ontologically equivalent and congruent.

Again, not so fast. The doctrine of genetic determinism is falsified by empirical evidence for epigenesis sensu lato (Vargas et al., 2020), rooted in the notion of living biological organisms as autopoietons (viz. autopoietic systems self-organized and maintained by hierarchically arranged systemic processes; Maturana & Varela, 1972; Maturana & Mpodozis, 2000; Mpodozis, 2022; cf. Hershkovitz, 2019b). In this framework, DNA basically "codes for" nothing but its own (imperfect) replication. Under *tolerable conditions*, however, the genome *can* interact with certain RNAs and proteins that are incidental to its milieu. It is these interactions that render the genome as a *metaphorical* template that does not "generate" RNAs and proteins so much as it *limits* the *sorts* of RNAs and proteins that, under tolerable conditions, *might be* generated. The RNAs and proteins, likewise, constrain linked downstream processes that eventually constrain the sorts of phenotypes/behaviors, and hence organismal species, that *might be* generated (viz. might *evolve*).

Contrary to reductionist doctrine, this epigenic[66] process works in *both* directions, viz. bottom-up and top-down. The phenotype/behavior of the organism is constrained by lower-level processes, but it is sui generis. It is *bound* to continue to behave/evolve (and systemically reproduce) as long as *tolerable* conditions permit. But at the same time, the form/behavior continuity of the organism also constrains form/behavior continuity at lower levels. If the organism dies, it dies at all levels. It also *influences* lower-level behavior/evolution. This is most obvious in the case of conventionally-conceived *molecular* epigenesis, e.g., methylation-regulated gene expression consequent to, e.g., "imprinting," and behaviorally-induced DNA transposition. Maturana & Mpodozis (2000; Mpodozis, 2022) thus referred to the *total* genome as comprising not merely the DNA genome, but the entire autopoietic system, viz. the organism and all of its material components (as well as its "ontogenetic niche"). The entire system (including physical and behavioral phenotypes) is inherited intergenerationally. Biological form/behavior/evolution is consequent to that of the entire system and not simply the DNA.

What does this have to do with phylogenomics? Phylogenomic evidence provides tantalizing evidence for the broad-sense notion of epigenesis and its axioms. Already conventional molecular systematic evidence more often than not has indicated polyphyly of DNA sequences associated with morphological species. This evidence usually is a function of the number of sampled individuals, which, historically, was "one." Commonly, researchers, including myself, have interpreted or restated this genotypic polyphyly as polyphyly of the *taxonomic* species themselves, but this sloppiness should be avoided unless *species* polyphyly itself is tested. Phylogenomic data indicates that genotypic polyphyly is even more common than appreciated. Are the associated taxa polyphyletic (and therefore convergent)?

In the epigenic framework, the question is whether the individuals maintained their taxonomic form/behavior *continuously* intergenerationally despite genomic alteration. In fact, we have known for decades that taxonomic species often manifest *macrogenomic* diversity, viz. karyotypic variation. This is not to say that genomic alteration does not alter also the organisms and their ontogenetic potential. But this is not the question. Rather, the question is whether the erstwhile ontogenetic phenotype that diagnoses the taxonomic species can continue more or less *unchanged* in the wake of DNA genomic modification. This would imply that species continuity is not dependent per se on continuity of its "genes," but rather merely continuity of its genetic *adequacy*.

Returning to *Leucheria*, theory and existing empirical evidence alone permit some predictions for the results. The computational output will include an "optimal" cladogram whose nodes are supported by

[66] Vargas et al. (2020) used the adjective "epigenic" rather than "epigenetic" to distinguish between epigenesis sensu lato and conventional molecular epigenesis.

bootstrap values mostly much higher than in the single-locus cladograms. Many cladogram nodes will be supported by [pseudo-]Bayesian PPs near unity. When analyzed separately, varying proportions of loci will support different relationships. These will not be "discoveries," because this reflects what we know about how genomes evolve and how the methods reconstruct this evolution. Going beyond this depends upon the researchers' theoretical and empirical knowledge and understanding, their abilities to frame specific questions, and their abilities apply methods sufficiently rigorously and creatively to answer these questions decisively and insightfully, without being misled by either dogma or analytical artifacts.

My own brief studies of *Leucheria*, in the context of my broader background, suggest to me dozens of specific questions, the methodological means of answering them, and the pitfalls to avoid. These questions go beyond that of what phylogenomic analysis might (correctly or incorrectly) divulge about *Leucheria* evolution. Just as important are questions of what *Leucheria* evolution divulges about genomic evolution, what this synthesis divulges about the nature of organisms and their evolution, and, ultimately, what scientific research divulges about the nature of science. But there is no point to articulating these specific questions and possible solutions here. My academic situation, or rather lack thereof, prohibits my undertaking of empirical phylogenomic studies. But at least I still can look forward to *correcting* them.

## Disclosure statement

The author declares no conflict of interest, viz. this research received no support from any politicized funding agency (e.g., government, NGO, or quasi-governmental international financial institution) or any so-funded research project or academic institution. The present work is entirely that of the author whose contribution to and responsibility for the work is 100%, whose authorship is not "vicarious" or editorially influenced by unacknowledged persons, and whose integrity is not potentially compromised for economic, social, or political expedience.

## Acknowledgments

Elaboration of my 38 publications since 2018 and many more to follow would not have been possible except for the generous support and dedicated efforts of Fundación Reshet in Chile (www.reshet.cl).

## Supplemental files

The separate ITS and cpDNA data sets in NEXUS format for this work is available at: https://files.osf.io/v1/resources/xs7dm/providers/osfstorage/?zip=

## Literature cited

Apodaca MJ, Palacio FX, Katinas L. Morphology and morphometry support a single species in the *Leucheria cerberoana* species complex (Asteraceae, Nassauvieae). Phytotaxa, 2021; 489(1) : 27-48. www.mapress.com/pt/article/view/phytotaxa.489.1.3

Archie JW. Homoplasy excess ratios: new indices for measuring levels of homoplasy in phylogenetic systematics and a critique of the consistency index. Systematic Biology, 1989; 38(3) : 253-269. https://doi.org/10.2307/2992286

Barker FK. Lutzoni FM. The utility of the incongruence length difference test. Systematic Biology, 2002; 51(4) : 625-637. https://doi.org/10.1080/10635150290102302

Böhnert T, Luebert F, Ritter B, Merklinger FF, Stoll A, Schneider JV, Quandt D, Weigend M. Origin and diversification of *Cristaria* (Malvaceae) parallel Andean orogeny and onset of hyperaridity in the Atacama Desert. Global and Planetary Change, 2019; 181 : 102992. https://doi.org/10.1016/j.gloplacha.2019.102992

Böhnert T, Luebert F, Merklinger FF, Harpke D, Stoll A, Schneider JV, Blattner FR, Quandt D, Weigend M. Plant migration under long-lasting hyperaridity – phylogenomics unravels recent biogeographic history in one of the oldest deserts on Earth. New Phytologist, 2022; 234(5) : 1863-1875. https://doi.org/10.1111/nph.18082

Bromham L, Duchêne S, Hua X, Ritchie AM, Duchêne DA, Ho SYW. Bayesian molecular dating: opening up the black box. Biological Review, 2018; 93(2) : 1165-1191. https://doi.org/10.1111/brv.12390

Candolle AP de. Prodromus systematis naturalis regni vegetabilis, volume 7. Paris: Sumptibus Sociorum Treuttel et Wurtz; 1838. https://www.biodiversitylibrary.org/item/7156

Cao R, Tong S, Luan T, Zheng H, Zhang W. Compensatory base changes and varying phylogenetic effects on angiosperm ITS2 genetic distances. Plants, 2022; 11(1) : 929. https://doi.org/10.3390/plants11070929

Crisci JV. Revisión del género *Leucheria* (Compositae: Mutisieae). Darwiniana, 1976; 20 : 9-126. https://www.jstor.org/stable/23215578

Delprete P, Forneris GG, Pistarino A. Carlo Bertero (1789–1831) in the New World. SIDA, 2002; 20 : 621-644. http://www.jstor.org/stable/41968082

Farris JS. The retention index and the rescaled consistency index. Cladistics, 1989; 5(4) : 417-419. https://doi.org/10.1111/j.1096-0031.1989.tb00573.x

Farris JS, Källersjö M, Kluge AG, Bult C. Testing significance of incongruence. Cladistics, 1994; 10(3) : 315-319. https://doi.org/10.1111/j.1096-0031.1994.tb00181.x

Farris JS, Källersjö M, Kluge AG, Bult C. Constructing a significance test for incongruence. Systematic Biology, 1995; 44(4) : 570-572. https://doi.org/10.2307/2413663

GBIF Secretariat. GBIF Backbone Taxonomy Checklist dataset. 2017. https://doi.org/10.15468/39omei

Goloboff PA. Calculating SPR distances between trees. Cladistics, 2008; 24(4) : 591-597. https://doi.org/10.1111/j.1096-0031.2007.00189.x

Gorban II. The Statistical Stability Phenomenon. Switzerland: Springer International Publishing; 2017. https://doi.org/10.1007/978-3-319-43585-5

Hershkovitz MA. Systematics, evolution, and phylogeography of Montiaceae (Portulacineae). Phytoneuron, 2019a; 2019-27 : 1–77. http://www.phytoneuron.net/2019Phytoneuron/27PhytoN-Montiaceae.pdf

Hershkovitz MA. The 'Holy Grail' in phylogenetic reconstruction: seeing the forest for the trees? EcoEvoRxiv, 2019b; B9MTN. https://doi.org/10.32942/osf.io/b9mtn

Hershkovitz M[A]. Bertero's ghost revisited: new typifications of *Talinum linaria* Colla and *Calandrinia gaudichaudii* Barnéoud (= *Calandrinia pilosiuscula* DC; Montiaceae). EcoEvoRxiv, 2020; N4D5J. https://doi.org/10.32942/osf.io/n4d5j

Hershkovitz MA. Evolutionary insights from DNA sequences from *Chaetanthera* Ruiz & Pav. and *Oriastrum* Poepp. & Endl. (Asteraceae; Mutisieae). I. Of molecules and systematics. EcoEvoRxiv, 2021a; AK68M. https://doi.org/10.32942/osf.io/ak68m/

Hershkovitz MA. Reconciliation of evidence for the Portulacineae "backbone" phylogeny. EcoEvoRxiv, 2021b; YRDSF. https://doi.org/10.32942/osf.io/yrdsf

Hershkovitz MA. CONFOUND IT! The taxonomy of plants mistaken for *Cistanthe arenaria* (Cham.) Carolin ex Hershk. [MONTIACEAE, *Cistantheae* sect. *Rosulatae* (Reiche) Hershk.], with notes on systematics and evolution of the genus *Cistanthe*. EcoEvoRxiv, 2022; X2G593. https://doi.org/10.32942/X2G593

Hershkovitz MA. Additions and Corrections to the Taxonomy of "*Cistanthe arenaria*" (Montiaceae). EcoEvoRxiv, 2023; X23K6P. https://doi.org/10.32942/X23K6P

Hershkovitz MA. *Leucheria* Lag. includes *Polyachyrus* Lag. (Asteraceae; Nassauvieae), International Journal of Science, Development and Research, 2024a; 9(4) : 636-646. https://doi.org/10.13140/RG.2.2.16344.71686/1

Hershkovitz MA. Revised molecular phylogenetic analysis of *Leucheria* Lag. sensu lato (Asteraceae; Nassauvieae) and implications for morphological and ecological evolution. EcoEvoRxiv, 2024b; X2NC9Q. https://doi.org/10.32942/X2NC9Q

Hershkovitz MA. Phylogenetic relations of *Macrachaenium* Hook.f. (Asteraceae; Mutisieae). EcoEvoRxiv, 2024c; X2XP5S. https://doi.org/10.32942/X2XP5S

Hershkovitz MA. A new name and two new combinations in *Leucheria* Lag. (Asteraceae; Nassauvieae). World Journal of Advanced Research and Reviews, 2024d; 22(1) : 764-765. https://doi.org/10.30574/wjarr.2024.22.1.1180

Hershkovitz MA. A phylogenetic sectional taxonomy of *Leucheria* Lag. (Asteraceae; Nassauvieae). World Journal of Advanced Research and Reviews, 2024; 24(3) : 1846-1849. https://doi.org/10.30574/wjarr.2024.24.3.3884

Hershkovitz MA. Phenological and life history/form idiosyncraticity of *Cistanthe philhershkovitziana* Hershk. (*C.* sect. *Cistanthe*; Montiaceae) in Chile's southern Central Littoral Zone. EcoEvoRxiv, 2024f; X2HS6G. https://doi.org/10.32942/X2HS6G.

Hershkovitz MA, Lewis LA. Deep-level diagnostic value of the rDNA-ITS region. Molecular Biology and Evolution, 1996; 13(9) : 1276-1295. https://doi.org/10.1093/oxfordjournals.molbev.a025693

Hershkovitz MA, Zimmer EA. Conservation patterns in angiosperm rDNA-ITS2 sequences. Nucleic Acids Research, 1996; 24(15) : 2857-2867. https://doi.org/10.1093/nar/24.15.2857

Huson DH, Bryant D. Application of phylogenetic networks in evolutionary studies. Molecular Biology and Evolution, 2006; 23(2) : 254-267. https://doi.org/10.1093/molbev/msj030

Jara-Arancio P, Vidal PM, Panero JL, Marticorena A, Arancio G, Arroyo MTK. Phylogenetic reconstruction of the South American genus *Leucheria* Lag. (Asteraceae, Nassauvieae) based on nuclear and chloroplast DNA sequences. Plant Systematics and Evolution, 2017; 303(2) : 221-232. https://doi.org/10.1007/s00606-016-1366-7

Jara-Arancio P, Vidal PM, Arroyo MTK. Phylogenetic reconstruction of the genus *Triptilion* (Asteraceae, Nassauvieae) based on nuclear and chloroplast DNA sequences. Journal of Systematics and Evolution, 2017b ["2018"]; 56(2) : 120-128. https://doi.org/10.1111/jse.12294

Jara-Arancio P, Ratto F, Bartoli A, Arancio G, Carmona-Ortiz MR. A new species of the genus *Leucheria* (Asteraceae, Nassauvieae) from Argentina. Phytotaxa, 2019; 404(1) : 51-57. URL

Katinas L, Apodaca MJ, Crisci JV. A synopsis of *Leucheria* (Asteraceae, Nassauvieae), with notes on the morphology. Washington DC, USA: Smithsonian Scholarly Press; 2022. https://notablesdelaciencia.conicet.gov.ar/bitstream/handle/11336/187978/CONICET_Digital_Nro.f6137c42-32ee-42c1-9103-f29ee8e0b549_B.pdf?sequence=2&isAllowed=y

Katinas L, Forte N. Capitulum compartmentalization in *Leucheria* (Nassauvieae): insights into the evolution of Asteraceae inflorescence. Taxon, 2020; 69(4) : 679-693. https://doi.org/10.1002/tax.12275

Katoh K, Misawa K, Kuma K, Miyata T. MAFFT: a novel method for rapid multiple sequence alignment based on fast Fourier transform. Nucleic Acids Research, 2002; 30(14) : 3059-3066. https://doi.org/10.1093/nar/gkf436

Lavandero N, Pérez [M]F, Pinilla N. *Leucheria peteroana* (Nassauvieae, Asteraceae), a new species of *Leucheria* endemic to the Andes of Central Chile, and insights into the systematics of Nassauviae [sic]. PhytoKeys, 2024; 248 : 315-337.

Lavandero N, Rosende B, Pérez MF. *Leucheria cantillanensis* (Nassauvieae, Asteraceae), a new species endemic to central Chile. PhytoKeys, 2020; 169 : 99-117. https://doi.org/10.3897/phytokeys.169.57532

Luebert F, Wen J, Dillon MO. Systematic placement and biogeographical relationships of the monotypic genera *Gypothamnium* and *Oxyphyllum* (Asteraceae: Mutisioideae) from the Atacama Desert. Botanical Journal of the Linnean Society, 2009; 159(1) : 32-51. https://doi.org/10.1111/j.1095-8339.2008.00926.x

Maturana HR, Mpodozis J. The origin of species by means of natural drift. Revista Chilena de Historia Natural, 2000; 73(2) : 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 HR, Varela FJ. Autopoiesis and cognition: the realization of the living. Boston studies in the philosophy and history of science, ed. 1. Dordrecht, Germany: Reidel; 1972.

Mpodozis J. Natural drift: a minimal theory with maximal consequences. Constructivist Foundations, 2022; 18(1) : 94-101. https://constructivist.info/18/1/094

Mickevich MF, Farris JS. The implications of congruence in Menidia. Systematic Zoology, 1981; 30(3) : 351-370. https://doi.org/10.2307/2413255

Nicola MV, Johnson LA, Pozner R. Unraveling patterns and processes of diversification in the South Andean-Patagonian *Nassauvia* subgenus *Strongyloma* (Asteraceae, Nassauvieae). Molecular Phylogenetics and Evolution, 2019; 136 : 164-182. https://doi.org/10.1016/j.ympev.2019.03.004

Pérez F, Lavandero N, Ossa CG, Hinojosa LF, Jara-Arancio P, Arroyo MTK. Divergence in plant traits and increased modularity underlie repeated transitions between low and high elevations in the Andean genus *Leucheria*. Frontiers in Plant Science, 2020; 11 : 714. https://www.frontiersin.org/article/10.3389/fpls.2020.00714

Planet PJ. Tree disagreement: measuring and testing incongruence in phylogenies. Journal of Biomedical Informatics, 2005 ["2006"]; 39(1) : 86-102. https://doi.org/10.1016/j.jbi.2005.08.008

Posada D, Buckley TR. Model selection and model averaging in phylogenetics: advantages of Akaike information criterion and Bayesian approaches over likelihood ratio tests. Systematic Biology, 2004; 53(5) : 793-808. https://doi.org/10.1080/10635150490522304

Qin Y, Li M, Cao Y, Gao Y, Zhang A. Molecular thresholds of ITS2 and their implications for molecular evolution and species identification in seed plants. Science Reports, 2017; 7(1) : 17316. https://doi.org/10.1038/s41598-017-17695-2

Reid NM, Carstens BC. Phylogenetic estimation error can decrease the accuracy of species delimitation: a Bayesian implementation of the general mixed Yule-coalescent model. BMC Evolutionary Biology, 2012; 12(1) : 196. https://doi.org/10.1186/1471-2148-12-196

Schwartz RS, Mueller RL. Branch length estimation and divergence dating: estimates of error in Bayesian and maximum likelihood frameworks. BMC Evolutionary Biology, 2010; 10 : 5. https://doi.org/10.1186/1471-2148-10-5

Swenson U, Lepschi B, Hutabarat PWK, Cronholm B, Nylinder S. Systematics of *Planchonella* Pierre: the largest Sapotaceae genus in subfamily Chrysophylloideae. Taxon, 2025. https://doi.org/10.1002/tax.13320

Swofford DL. PAUP*. Phylogenetic Analysis Using Parsimony (*and Other Methods). Version 4. Sunderland, Massachusetts USA: Sinauer Associates; 2003. https://paup.phylosolutions.com/

Swofford DL, Olsen GJ, Waddell PJ, Hillis PM. Phylogenetic inference. Pp. 407-514 in DM Hillis, C Moritz, BK Mable (editors), Molecular Systematics, 2nd ed. Sunderland, Massachusetts USA: Sinauer Associates; 1996.

Thompson JD, Gibson TJ, Higgins DG. Multiple sequence alignment using ClustalW and ClustalX. Current Protocols in Bioinformatics, 2002; 2.3.1-2.3.22. https://doi.org/10.1002/0471250953.bi0203s00

Vargas AO, Botelho JF, Mpodozis J. The evolutionary consequences of epigenesis and neutral change: A conceptual approach at the organismal level. Journal of Experimental Zoology. Part B: Molecular and Developmental Evolution, 2023; 340(8) : 531-540. https://doi.org/10.1002/jez.b.23023

Wang W, Li H, Chen Z. Analysis of plastid and nuclear DNA data in plant phylogenetics—evaluation and improvement. Science China Life Sciences, 2014; 57(3) : 280-286. https://doi.org/10.1007/s11427-014-4620-7

Williams DM, Ebach C. Cladistics, a Guide to Biological Classification, ed. 3. Cambridge, UK: Cambridge University Press; 2020. https://doi.org/10.1017/9781139047678

Wolf M, Chen S, Song J, Ankenbrand M, Müller T. Compensatory base changes in ITS2 secondary structures correlate with the biological species concept despite intragenomic variability in ITS2 sequences – a proof of concept. PLoS ONE, 2013; 8(6) : e66726. https://doi:10.1371/journal.pone.0066726

Yang Z, Goldman N, Friday A. Maximum likelihood trees from DNA sequences: a peculiar statistical estimation problem. Systematic Biology, 1995; 44(3) : 384-399. https://doi.org/10.1093/sysbio/44.3.384

**Figure captions**

**Figure 1.** Portion of the nrDNA ITS region alignment for several *Leucheria* species showing the 3' end of the 5.8S gene and the 5' end of ITS2 (beginning at position 441). The Jara2017 *L. diemii* var. *diemii* label is highlighted to draw attention to the large number of differences between this sequence and those of other species. Note that, in those positions where the *L. diemii* var. *diemii* sequence differs, the other sequences are completely conserved.

**Figures 2–6.** Cladograms/phylograms for the Leucheria crown group. Shaded boxes indicate *Leucheria* sections. Sequences (Appendix 1) are those used in Hershkovitz (2024b; *Leucheria* sequences from Jara2017) plus those from Lavandero2024 (denoted with an "L" following the taxon acronym) and an additional available sequence denoted with an "M." Boxed acronyms denote samples with identical sequences, ignoring (in some cases) probably spurious uninformative indels (see text). Acronyms in quotation marks refer to questioned or incorrect identifications (see text).

**Figure 2.** Strict consensus of 144 MP trees [rescaled consistency index (RC [Farris, 1989] = 0.56; homoplasy index (HI [1 – HER; Archie, 1989] = 0.37) for the ITS data excluding indel characters. The branches indicate MP bootstrap support for analyses excluding/including indel characters. Shaded bootstrap values indicate branches in conflict with the MP strict consensus including indel characters (Fig. 3). GC% of MP informative sites is indicated. The "eriocephala L" and "purpurea" sequences were inadvertently excluded from the analysis, and their position is here inserted manually. The sequences are identical to the "eriocephala" and "purpurea L" sequences, except that the latter each include a single autapomorphic substitution (see Fig. 4).

**Figure 3.** As in Fig. 1, but strict consensus of 679 MP trees (RC = 0.63; HI = 0.37) for the ITS data including indel characters. The figure highlights only topological and bootstrap differences from Fig. 1.

**Figure 4.** ML phylogram for the *Leucheria* crown group ITS data. The "eriocephala L" and "purpurea" sequences were inadvertently excluded from the analysis, and their position is here inserted manually. The sequences are identical to the "eriocephala" and "purpurea L" sequences, except that the latter each include a single autapomorphic substitution.

**Figure 5.** Strict consensus of 26,988 MP trees (RC = 0.89; HI = 0.08) for the cpDNA data excluding indel characters (left side), with an inset showing a portion of the MP strict consensus tree including indels (RC = 0.86; HI = 0.11). The topological differences of the latter are limited to *L.* sect. *Cassiopea*. The branches indicate MP bootstrap support for analyses excluding/including indel characters. Shaded bootstrap values indicate branches in conflict with the complementary MP strict consensus with the contrary indel inclusion criterion. GC% of MP informative sites is indicated.

**Figure 6.** ML phylogram for the *Leucheria* crown group cpDNA data.

**Figure 7.** Portion of the *trnL-trnF* alignment showing the position and distribution of a 19bp indel and adjacent (but independent) 6bp indel. Both indels are homoplasious.

**Figure 1**

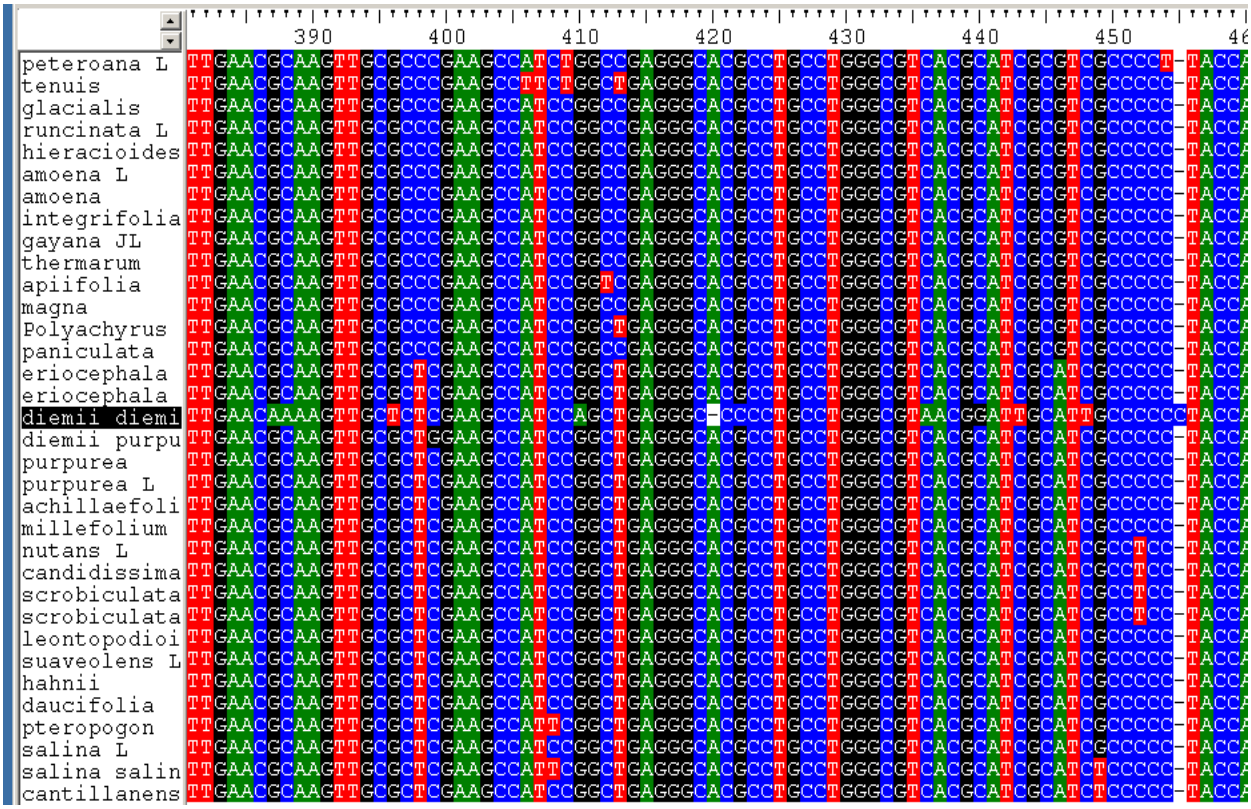

390
400
410
420
430
440
450
peteroana L
tenuis
glacialis
runcinata L
hieracioides
amoena L
amoena
integrifolia
gayana JL
thermarum
apiifolia
magna
Polyachyrus
paniculata
eriocephala
eriocephala
diemii diemi
diemii purpu
purpurea
purpurea L
achillaefoli
millefolium
nutans L
candidissima
scrobiculata
scrobiculata
leontopodioi
suaveolens L
hahnii
daucifolia
pteropogon
salina L
salina salin
cantillanens

**Figure 2**

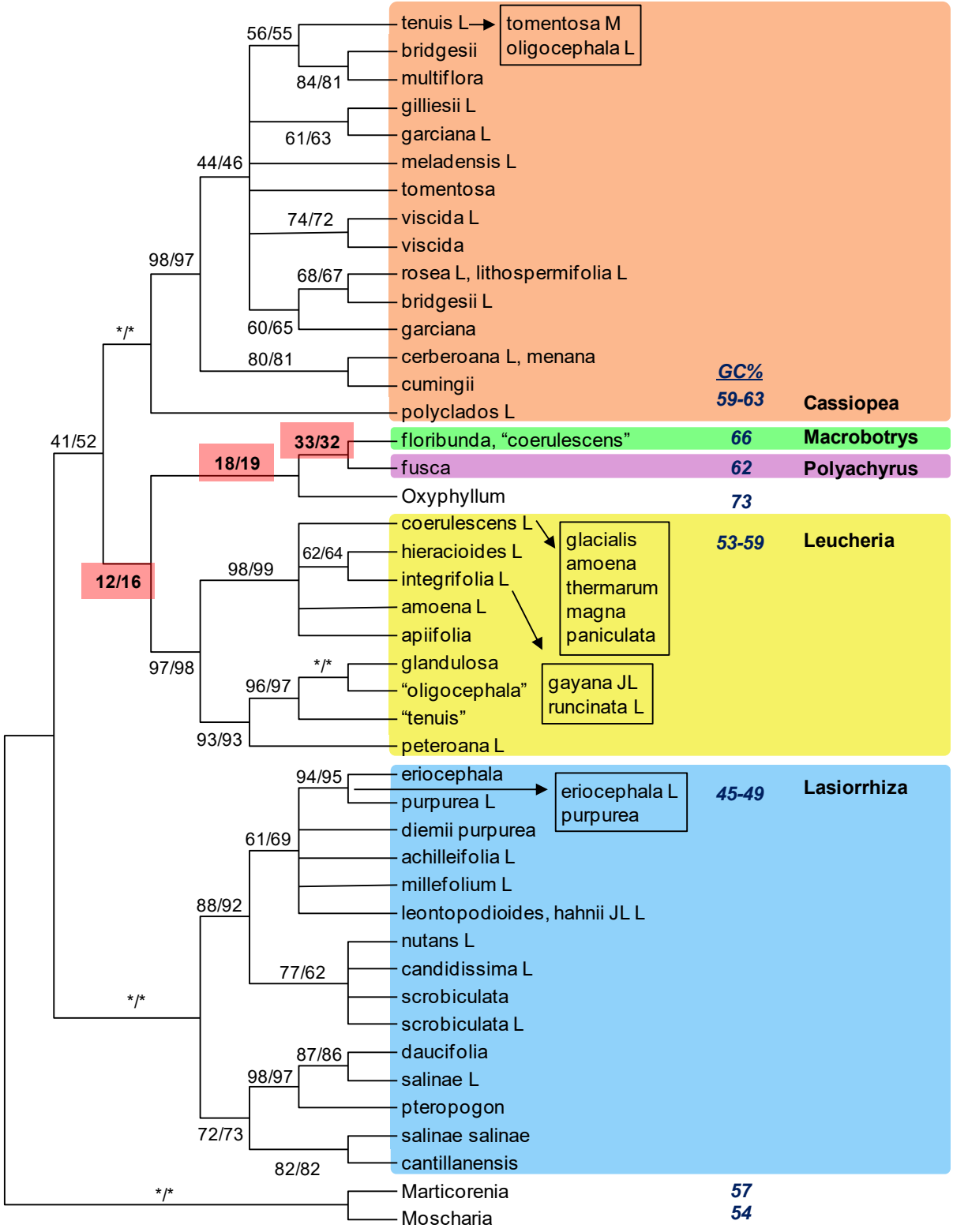

tenuis L
tomentosa M
oligocephala L
bridgesii
multiflora
gilliesii L
garciana L
meladensis L
tomentosa
viscida L
viscida
rosea L, lithospermifolia L
bridgesii L
garciana
cerberoana L, menana
cumingii
polyclados L
56/55
84/81
61/63
44/46
74/72
98/97
68/67
60/65
80/81
*/*
GC%
59-63
Cassiopea
floribunda, "coerulescens"
66
Macrobotrys
fusca
62
Polyachyrus
Oxyphyllum
73
41/52
33/32
18/19
12/16
coerulescens L
hieracioides L
integrifolia L
amoena L
apiifolia
glandulosa
"oligocephala"
"tenuis"
peteroana L
glacialis
amoena
thermarum
magna
paniculata
gayana JL
runcinata L
53-59
Leucheria
62/64
98/99
97/98
*/*
96/97
93/93
eriocephala
purpurea L
diemii purpurea
achilleifolia L
millefolium L
leontopodioides, hahnii JL L
nutans L
candidissima L
scrobiculata
scrobiculata L
daucifolia
salinae L
pteropogon
salinae salinae
cantillanensis
eriocephala L
purpurea
45-49
Lasiorrhiza
94/95
61/69
88/92
77/62
*/*
87/86
98/97
72/73
82/82
*/*
Marticorenia
Moscharia
57
54

**Figure 3**

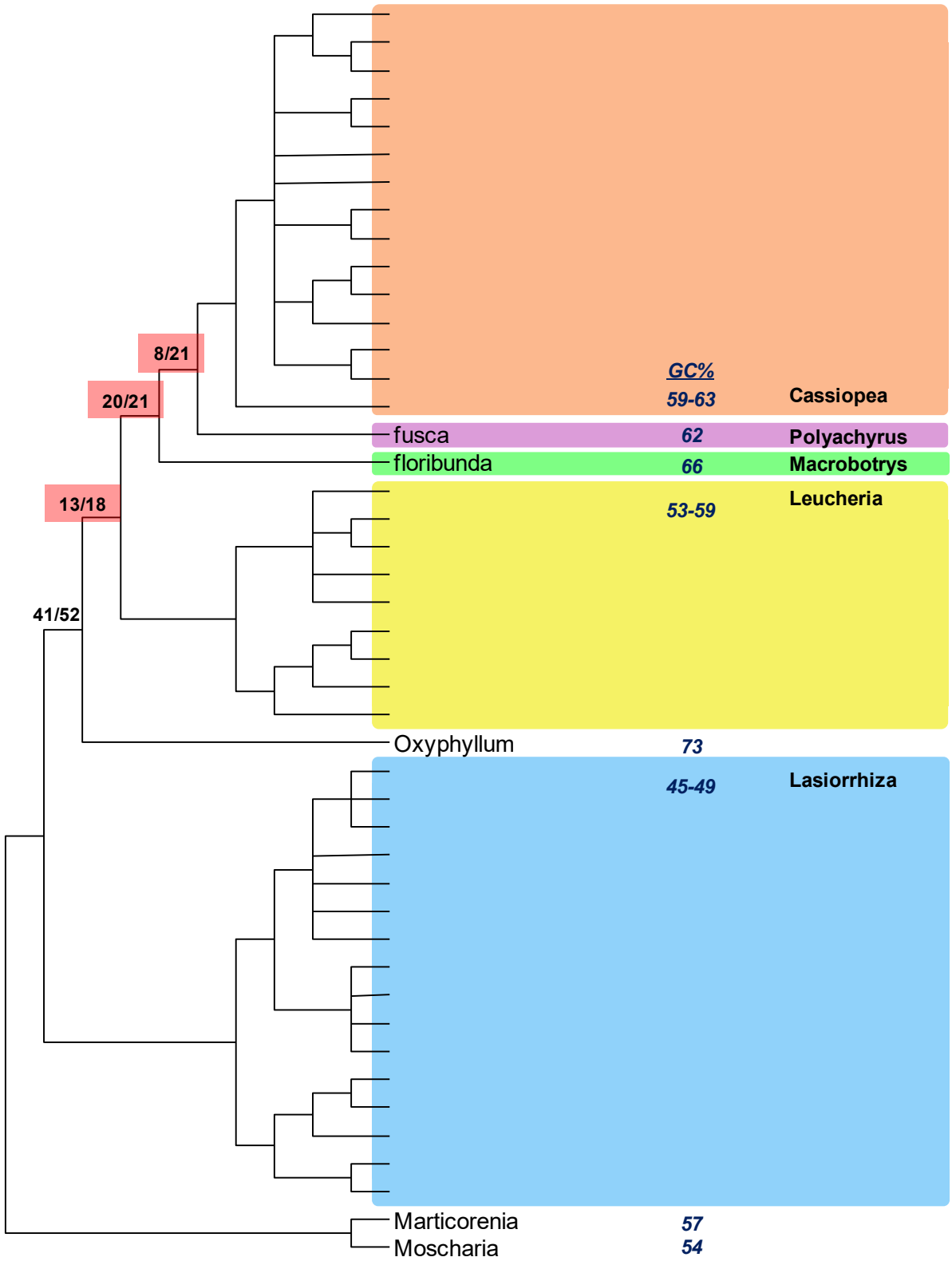

8/21
20/21
13/18
41/52
GC%
59-63
Cassiopea
fusca
62
Polyachyrus
floribunda
66
Macrobotrys
53-59
Leucheria
Oxyphyllum
73
45-49
Lasiorrhiza
Marticorenia
57
Moscharia
54

**Figure 4**

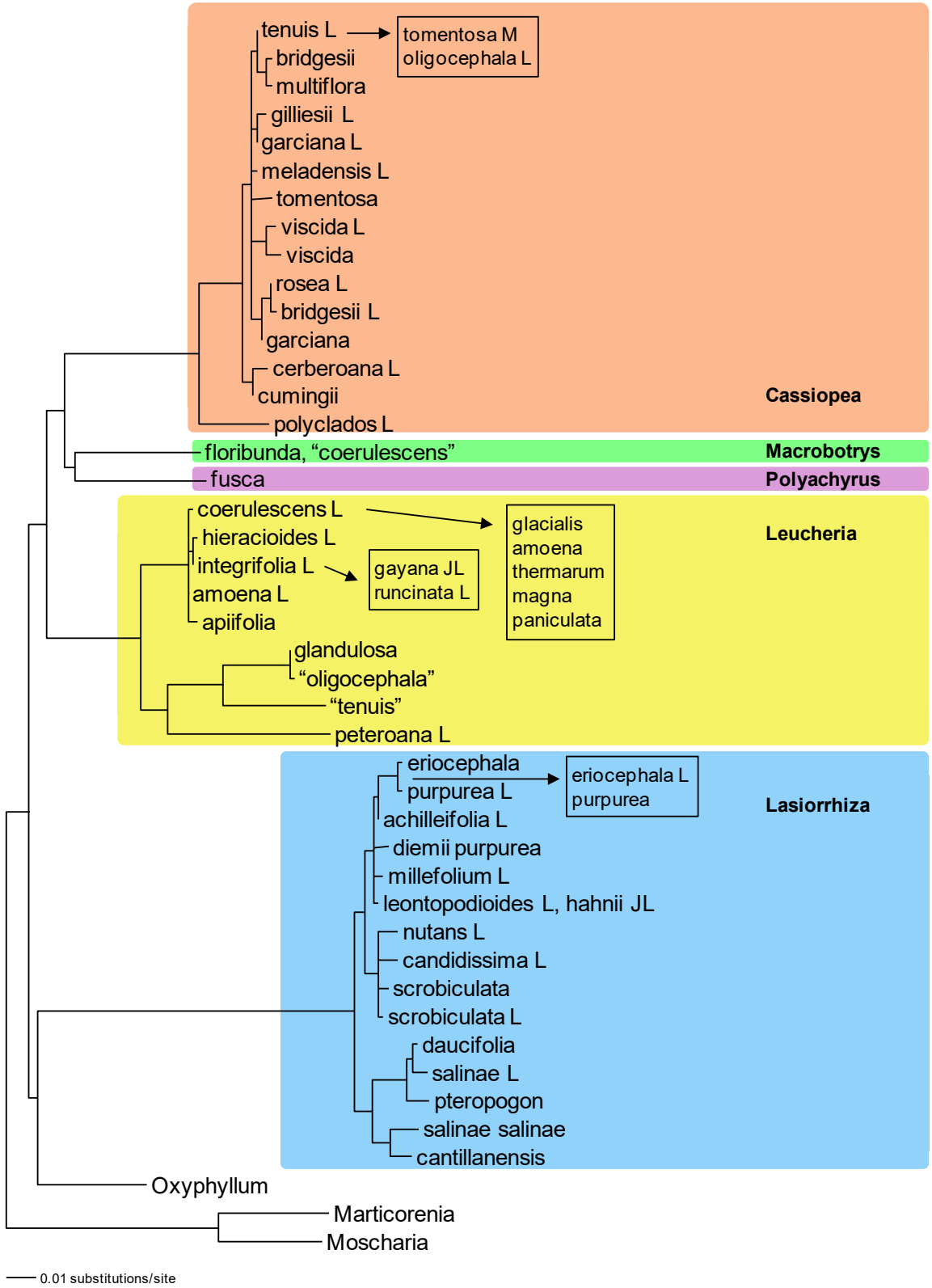

tenuis L
tomentosa M
oligocephala L
bridgesii
multiflora
gilliesii L
garciana L
meladensis L
tomentosa
viscida L
viscida
rosea L
bridgesii L
garciana
cerberoana L
cumingii
polyclados L
Cassiopea
floribunda, "coerulescens"
Macrobotrys
fusca
Polyachyrus
coerulescens L
hieracioides L
integrifolia L
amoena L
apiifolia
gayana JL
runcinata L
glacialis
amoena
thermarum
magna
paniculata
Leucheria
glandulosa
"oligocephala"
"tenuis"
peteroana L
eriocephala
purpurea L
eriocephala L
purpurea
achilleifolia L
diemii purpurea
millefolium L
leontopodioides L, hahnii JL
nutans L
candidissima L
scrobiculata
scrobiculata L
daucifolia
salinae L
pteropogon
salinae salinae
cantillanensis
Lasiorrhiza
Oxyphyllum
Marticorenia
Moscharia
0.01 substitutions/site

**Figure 5**

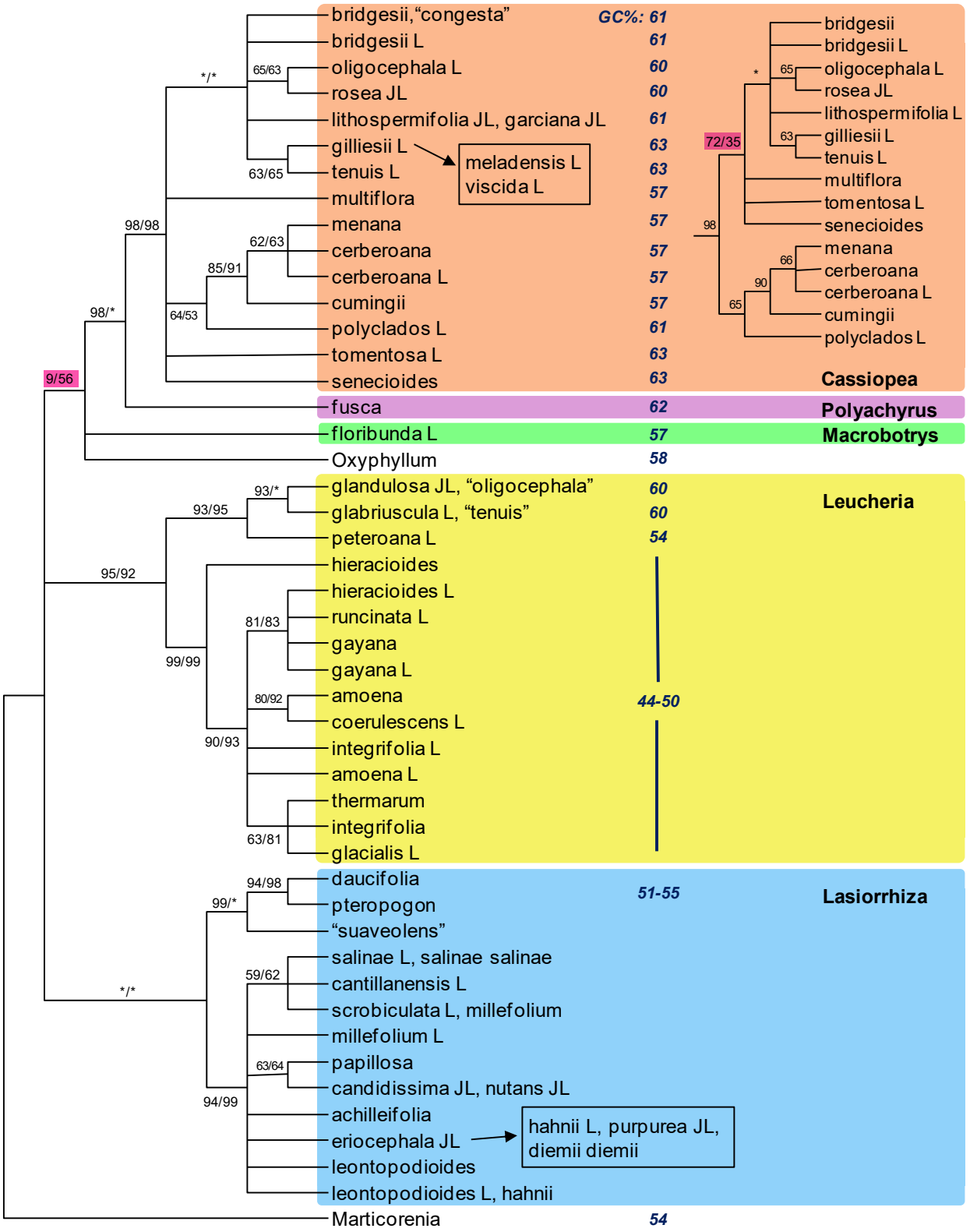

bridgesii,"congesta"
bridgesii L
oligocephala L
rosea JL
lithospermifolia JL, garciana JL
gilliesii L
tenuis L
multiflora
menana
cerberoana
cerberoana L
cumingii
polyclados L
tomentosa L
senecioides
meladensis L
viscida L
GC%: 61
61
60
60
61
63
63
57
57
57
57
61
63
63
bridgesii
bridgesii L
oligocephala L
rosea JL
lithospermifolia L
gilliesii L
tenuis L
multiflora
tomentosa L
senecioides
menana
cerberoana
cerberoana L
cumingii
polyclados L
Cassiopea
fusca
62
Polyachyrus
floribunda L
57
Macrobotrys
Oxyphyllum
58
glandulosa JL, "oligocephala"
glabriuscula L, "tenuis"
peteroana L
hieracioides
hieracioides L
runcinata L
gayana
gayana L
amoena
coerulescens L
integrifolia L
amoena L
thermarum
integrifolia
glacialis L
60
60
54
44-50
Leucheria
daucifolia
pteropogon
"suaveolens"
salinae L, salinae salinae
cantillanensis L
scrobiculata L, millefolium
millefolium L
papillosa
candidissima JL, nutans JL
achilleifolia
eriocephala JL
leontopodioides
leontopodioides L, hahnii
hahnii L, purpurea JL, diemii diemii
51-55
Lasiorrhiza
Marticorenia
54
9/56
98/98
98/*
*/*
65/63
63/65
62/63
85/91
64/53
72/35
98
*
65
63
66
90
65
93/*
93/95
95/92
99/99
81/83
80/92
90/93
63/81
94/98
99/*
59/62
*/*
63/64
94/99

**Figure 6**

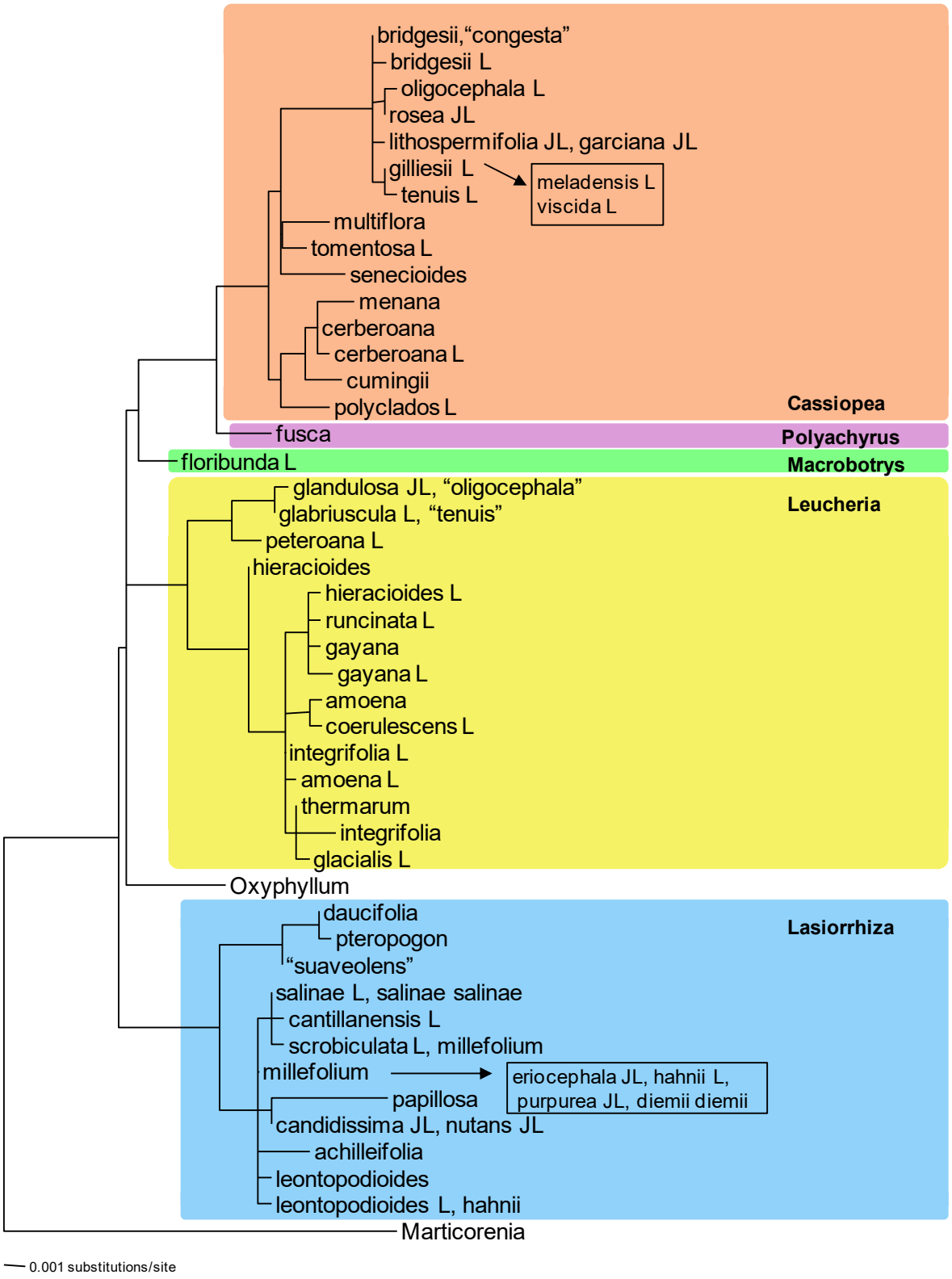
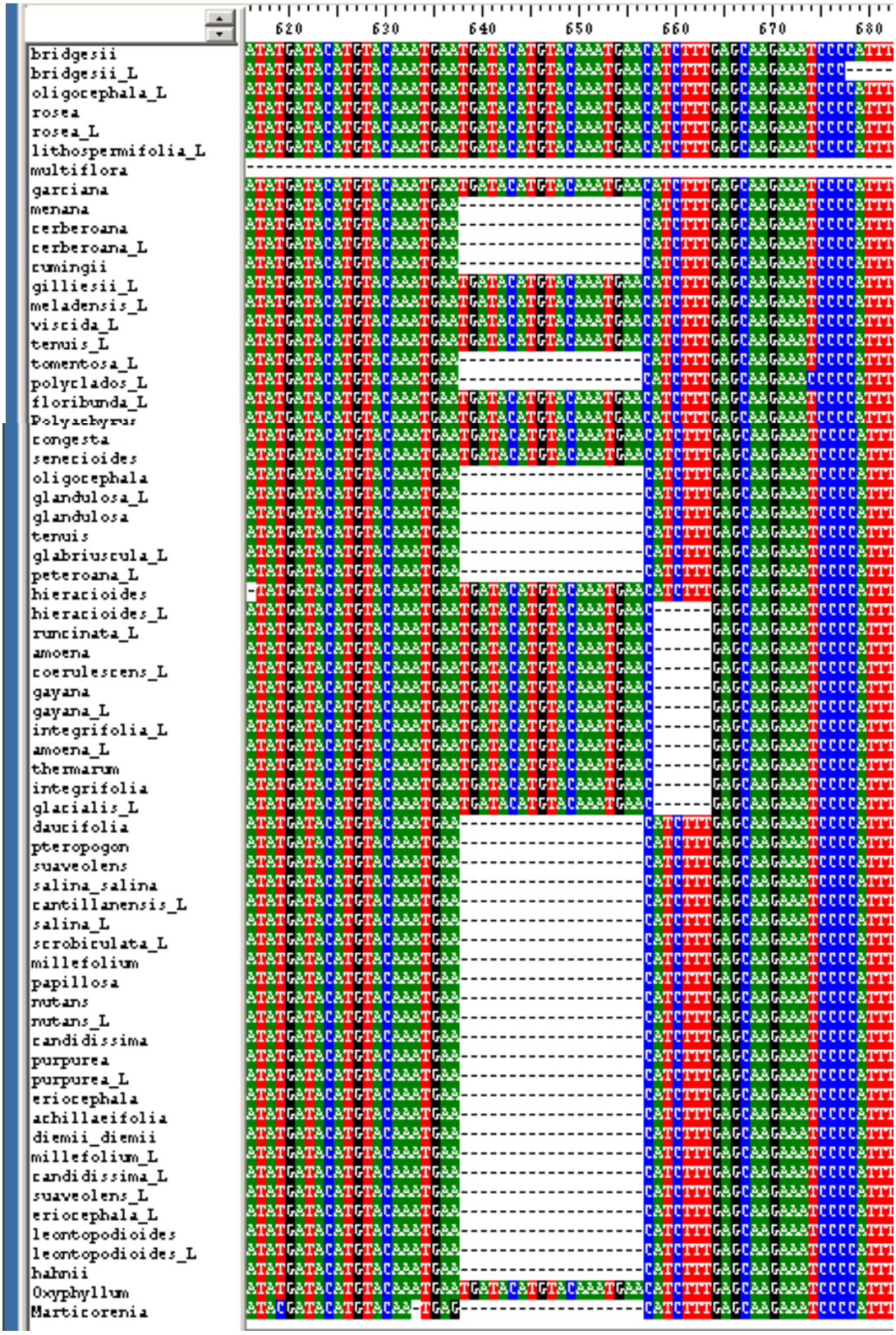

bridgesii,"congesta"
bridgesii L
oligocephala L
rosea JL
lithospermifolia JL, garciana JL
gilliesii L
tenuis L
meladensis L
viscida L
multiflora
tomentosa L
senecioides
menana
cerberoana
cerberoana L
cumingii
polyclados L
Cassiopea
fusca
Polyachyrus
floribunda L
Macrobotrys
glandulosa JL, "oligocephala"
glabriuscula L, "tenuis"
peteroana L
hieracioides
hieracioides L
runcinata L
gayana
gayana L
amoena
coerulescens L
integrifolia L
amoena L
thermarum
integrifolia
glacialis L
Leucheria
Oxyphyllum
daucifolia
pteropogon
"suaveolens"
salinae L, salinae salinae
cantillanensis L
scrobiculata L, millefolium
millefolium
eriocephala JL, hahnii L,
purpurea JL, diemii diemii
papillosa
candidissima JL, nutans JL
achilleifolia
leontopodioides
leontopodioides L, hahnii
Lasiorrhiza
Marticorenia
0.001 substitutions/site

**Figure 7**


620
630
640
650
660
670
680
bridgesii
bridgesii_L
oligocephala_L
rosea
rosea_L
lithospermifolia_L
multiflora
garciana
menana
cerberoana
cerberoana_L
cumingii
gilliesii_L
meladensis_L
viscida_L
tenuis_L
tomentosa_L
polyclados_L
floribunda_L
Polyachyrus
congesta
senecioides
oligocephala
glandulosa_L
glandulosa
tenuis
glabriuscula_L
peteroana_L
hieracioides
hieracioides_L
runcinata_L
amoena
coerulescens_L
gayana
gayana_L
integrifolia_L
amoena_L
thermarum
integrifolia
glacialis_L
dauciifolia
pteropogon
suaveolens
salina_salina
cantillanensis_L
salina_L
scrobiculata_L
millefolium
papillosa
nutans
nutans_L
candidissima
purpurea
purpurea_L
eriocephala
achillaeifolia
diemii_diemii
millefolium_L
candidissima_L
suaveolens_L
eriocephala_L
leontopodioides
leontopodioides_L
hahnii
Oxyphyllum
Marticorenia

**Appendix 1.** Taxa sampled, their acronyms in Figures 2–6, and their Genbank, ITS, *trnL-trnF*, and *rpl32-trnL* sequence. The Genbank sequences were trimmed at the 5' and/or 3' ends to fit the alignments constructed here. Questioned or incorrect taxon identifications are indicated by quotation marks. The acronyms followed by the letter "L" denote sequences used in Lavandero2024. Those without a letter or with "[J]" denote sequences used in Jara2017 (and Lavandero2020). The "J" is included in figures where the sequences are (actually or effectively) identical to the "L" sequences. One acronym is followed by the letter "M," denoting an additional available Genbank sequence.

| Genus | species | Acronym | ITS | rpl32-trnL | trnL-trnF |
|---|---|---|---|---|---|
| *Leucheria* | *achilleifolia*[67] | achilleifolia L | PQ059440 | – | – |
| *Leucheria* | *achilleifolia* | achilleifolia | – | KY223757 | KY223808 |
| *Leucheria* | *amoena* | amoena L | PQ059441 | PQ120809 | PQ120775 |
| *Leucheria* | *amoena* | amoena | KY010387 | KY223762 | KY223814 |
| *Leucheria* | *apiifolia* | apiifolia | KY010392 | – | |
| *Leucheria* | *bridgesii* | bridgesii L | PQ059442 | PQ120819 | PQ120782 |
| *Leucheria* | *bridgesii* | bridgesii | KY010363 | KY223742 | KY223791 |
| *Leucheria* | *candidissima* | candidissima L | PQ059443 | PQ120799 | PQ120771 |
| *Leucheria* | *candidissima* | candidissima [J] | – | KY223773 | KY223824 |
| *Leucheria* | *cantillanensis* | cantillanensis L | MT425194 | MT419957 | MT419958 |
| *Leucheria* | *cerberoana* | cerberoana L | PQ059444 | PQ120828 | PQ120789 |
| *Leucheria* | *cerberoana* | cerberoana | – | KY223769 | KY223820 |
| *Leucheria* | *coerulescens* | coerulescens L | PQ059445 | PQ120814 | PQ120780 |
| *Leucheria* | "*coerulescens*" | "coerulescens" | KY010377 | – | – |
| *Leucheria* | *"congesta*" | "congesta" | – | KY223743 | KY223792 |
| *Leucheria* | *cumingii* | cumingii | KY010404 | KY223778 | KY223830 |
| *Leucheria* | *daucifolia* | daucifolia | KY010398 | KY223772 | KY223823 |
| *Leucheria* | *diemii var. diemii* | diemii diemii | KY010402 | KY223776 | KY223827 |
| *Leucheria* | *diemii var. purpurea* | diemii purpurea | KY010403 | – | |
| *Leucheria* | *eriocephala* | eriocephala L | PQ059446 | PQ120800 | PQ120766 |
| *Leucheria* | *eriocephala* | eriocephala [J] | KY010365 | KY223744 | KY223793 |
| *Leucheria* | *floribunda* | floribunda L | – | PQ120808 | PQ120793 |
| *Leucheria* | *floribunda* | floribunda | KY010382 | – | – |
| *Leucheria* | *fusca* | fusca | EF530261 | MG553855 | EF530309 |
| *Leucheria* | *garciana* | garciana L | PQ059448 | PQ120820 | PQ120790 |
| *Leucheria* | *garciana* | garciana [J] | KY010401 | KY223775 | KY223826 |
| *Leucheria* | *gayana* | gayana L | PQ059449 | PQ120815 | PQ120776 |
| *Leucheria* | *gayana* | gayana [J] | KY009611 | KY223741 | KY223790 |
| *Leucheria* | *gilliesii* | gilliesii L | PQ059450 | PQ120823 | PQ120783 |
| *Leucheria* | *glabriuscula* | glabriuscula L | – | PQ120817 | PQ120795 |

[67] Katinas2022, among others (including Jara2017 and Hershkovitz, 2024b) used the incorrect spelling "achllaeifolia" from the original publication. But the epithet is based on *Achillea* L.

| | | | | | |
|---|---|---|---|---|---|
| *Leucheria* | *glacialis* | glacialis L | PQ059451 | – | – |
| *Leucheria* | *glacialis* | glacialis | KY010391 | KY223765 | KY223817 |
| *Leucheria* | *glandulosa* | glandulosa L | – | PQ120818 | PQ120796 |
| *Leucheria* | *glandulosa* | glandulosa | KY010366 | KY223745 | KY223794 |
| *Leucheria* | *hahnii*[68] | hahnii L | – | PQ120801 | PQ120767 |
| *Leucheria* | *hahnii* | hahnii [J] | KY010394 | KY223768 | KY223819 |
| *Leucheria* | *hieracioides* | hieracioides L | PQ059453 | PQ120811 | PQ120781 |
| *Leucheria* | *hieracioides* | hieracioides | – | KY223746 | KY223795 |
| *Leucheria* | *integrifolia*[69] | integrifolia L | PQ059454 | PQ120810 | PQ120778 |
| *Leucheria* | *integrifolia* | integrifolia | KY010393 | KY223767 | KY223818 |
| *Leucheria* | *leontopodioides* | leontopodioides L | PQ059455 | PQ120807 | PQ120768 |
| *Leucheria* | *leontopodioides* | leontopodioides | – | KY223759 | KY223810 |
| *Leucheria* | *lithospermifolia* | lithospermifolia L | PQ059456 | PQ120821 | PQ120791 |
| *Leucheria* | *lithospermifolia* | lithospermifolia [J] | KY010400 | KY223774 | KY223825 |
| *Leucheria* | *magna* | magna | KY010389 | – | – |
| *Leucheria* | *meladensis* | meladensis L | PQ059457 | PQ120824 | PQ120784 |
| *Leucheria* | *menana* | menana | KY010378 | KY223755 | KY223806 |
| *Leucheria* | *millefolium* | millefolium L | PQ059458 | PQ120802 | PQ120769 |
| *Leucheria* | *millefolium* | millefolium | – | KY223764 | KY223816 |
| *Leucheria* | *multiflora* | multiflora | KY010368 | KY223796 | – |
| *Leucheria* | *nutans* | nutans L | PQ059459 | PQ120803 | PQ120772 |
| *Leucheria* | *nutans* | nutans[J] | – | KY223771 | KY223822 |
| *Leucheria* | *oligocephala* | oligocephala L | PQ059460 | PQ120827 | PQ120787 |
| *Leucheria* | "*oligocephala*" | "oligocephala" | KY010384 | KY223760 | KY223811 |
| *Leucheria* | *paniculata* | paniculata | KY010379 | – | – |
| *Leucheria* | *papillosa* | papillosa | KY010369 | KY223747 | KY223797 |
| *Leucheria* | *peteroana* | peteroana L | PQ059461 | PQ120816 | PQ120797 |
| *Leucheria* | *polyclados* | polyclados L | PQ059462 | PQ120829 | PQ120794 |
| *Leucheria* | *pteropogon* | pteropogon | KY010371 | KY223748 | KY223799 |
| *Leucheria* | *purpurea* | purpurea L | PQ059463 | PQ120804 | PQ120770 |
| *Leucheria* | *purpurea* | purpurea [J] | KY010372 | KY223749 | KY223800 |
| *Leucheria* | *rosea* | rosea L | PQ059464 | PQ120822 | PQ120788 |
| *Leucheria* | *rosea* | rosea [J] | – | KY223750 | KY223801 |
| *Leucheria* | *runcinata* | runcinata L | PQ059465 | PQ120812 | PQ120779 |

**68** Genbank classifies *L. hahnii* as *L. suaveolens* (evidently per Katinas2022). These sequences were submitted as *L. hahnii*, because they derive from continental Patagonian rather than Falklands insular collections.

**69** Genbank erroneously classifies *Leucheria integrifolia* in *L. suaveolens*. This evidently is consequent to a processing error in which the Lavandero2024 sequences of *L. hahnii* and *L. integrifolia* were switched. As noted, Genbank classifies *L. hahnii* as *L. suaveolens* (evidently per Katinas2022). Genbank later partially corrected the error, but in the process also merged *L. integrifolia* and *L. suaveolens*.

| | | | | | |
|---|---|---|---|---|---|
| *Leucheria* | *salinae* | salinae L | PQ059466 | PQ120805 | PQ120773 |
| *Leucheria* | *salinae ssp. salinae* | salinae salinae | KY010375 | KY223752 | KY223803 |
| *Leucheria* | *scrobiculata* | scrobiculata L | PQ059467 | PQ120806 | PQ120774 |
| *Leucheria* | *scrobiculata* | scrobiculata | KY010385 | – | – |
| *Leucheria* | *senecioides* | senecioides | – | KY223753 | KY223804 |
| *Leucheria* | "*suaveolens*" | "suaveolens" | – | KY223777 | KY223829 |
| *Leucheria* | *tenuis* | tenuis L | PQ059468 | PQ120826 | PQ120785 |
| *Leucheria* | "*tenuis*" | "tenuis" | KY006586 | – | – |
| *Leucheria* | *thermarum* | thermarum | KY010396 | KY223770 | KY223821 |
| *Leucheria* | *tomentosa* | tomentosa L | – | PQ120830 | PQ120792 |
| *Leucheria* | *tomentosa* | tomentosa M | MG553776 | – | – |
| *Leucheria* | *tomentosa* | tomentosa | KY010380 | – | – |
| *Leucheria* | *viscida* | viscida L | PQ059470 | PQ120825 | PQ120786 |
| *Leucheria* | *viscida* | viscida | KY010386 | – | |
| *Marticorenia* | *foliosa* | Marticorenia | KY006584 | KY223780 | KY223733 |
| *Moscharia* | *pinnatifida* | Moscharia | EF530218 | – | – |
| *Oxyphyllum* | *ulicinum* | Oxyphyllum | EU729344 | MG553854 | EU729340 |