Review of "Missing data, clade support, and 'reticulation:' the molecular systematics of Heliconius and related genera (Lepidoptera: Nymphalidae) reexamined by Brower and Garzón-Orduna

Oh dear, where to start! It would perhaps be more honest to have titled this paper: "A critique of the findings of reticulation in the paper by Kozak et al. 2015 (K15), and a general tirade against model-based phylogenetics and in favor of a parsimony analysis."

It is worth noting that Brower and Garzón-Orduna (hereafter B16) term reticulation a "non-phylogenetic" or "contra-phylogenetic" process; apparently, according to the authors, the true phylogeny may only involve bifurcation, even if, perhaps, true reticulation was involved (pattern cladistics). Thus even the abstract displays a bias against reticulation.

Another word "homoplasy" is used in the abstract as a process-based explanatory term to mean "evolutionary convergence" rather than just a pattern in the character states. Again, given that the authors apparently believe that evolution only involves strict bifurcation, this demonstrates bias.

The paper seems characterized by little bouts of bad temper and adjectives expressing distaste for the findings of the K15 paper. For example, in presenting the results of K15, B16 state:

"The most recent of these, generated from a 22-gene matrix by K15, presents, as its preferred hypothesis, a topology that is largely similar to those of Brower and Egan (1997) and Beltrán et al. (2007) (Fig. 1). What sets K15 apart as more than simply a corroboration of prior phylogenetic studies is its emphasis on finding 'pervasive conflict between the loci' (K15, p. 512) as inferred from results of a variety of multispecies coalescent analyses." Needless to say, Brower is an author on both the earlier papers for which he sees K15 as largely a "corroboration." In fact, K15 states several important aims that sets it well apart from those of the earlier papers, aims that are very different from those in B16's mischaracterization:

1) To place "nearly all of the currently valid species in the tribe" in a phylogenetic analysis

2) To apply a wide range of phylogenetic methods, particularly coalescent and network approaches that have not been previously considered.

3) To investigate marker heterogeneity, and (probabilistic) support values for the topologies recovered.

4) To estimate the times of divergence of the species while taking account of the gene-tree species-tree problem.

5) To analyze evidence for temporal heterogeneity of diversification in the group, leading to the conclusion that Heliconius is indeed a "radiation" that diversified more rapidly than related genera.

6) Although not cited in "Aims of the Study," K15 also tests (and rejects as producing nonsensical results) a phylogenetic model of species delimitation that tends to lump some known separate species, as well as splitting into species populations generally considered conspecific.

Only K15 findings on aim 3 are disputed by this scathing critique of B16.

Using the B16 authors' logic, because they find the same topology as in K15 using a low-powered, error-prone parsimony analysis, perhaps this whole paper should be argued to be merely a weak corroboration of K15's result.

Instead B16 concentrate on three main problems they see with the K15 analysis:

1) Missing data in a number of taxa: B16 regard these missing data as too extensive to allow much confidence in the topological placement of the low-coverage taxa. They use a non-probabilistic, non-model-based argument to obtain this conclusion, instead relying on their preferred parsimony-based "support" measures.

2) The use of reticulation to explain phylogenetic discordance among partitions of the data. B16 argue that it can all be explained in a bifurcation context as due to "homoplasy" (see above for the process-based definition they presumably intend) or incomplete lineage sorting.

3) Model-based statistical analysis. B16 regard parsimony-based approaches and values of support to be superior to probabilistic modeling (e.g. lines 153-170 on page 9).

In the end, I'm convinced by none of these arguments, and in in any case they in no way invalidate the main findings of K15. The angry tone of the B16 article seems to be unnecessary. Taking each of the major conclusions of B16 in order:

1) I'm not convinced that arguing from an anti-probabilistic parsimony approach, and parsimony support (particularly anti-Bayesian) point of view is very convincing evidence that missing data is a problem, as made out by B16. Personally, I'd prefer to see those species with much missing data to be placed in the phylogenetic hypothesis, and with some sort of measure of statistical probability in terms of support (for example bootstrap value, or Bayesian posterior). Bremer and other parsimony support values suggested do not provide this information. If the main purpose is to rail against modern statistical analyses, B16 must realize that they could have raised this particular argument against most of the other papers in Systematic Biology as well.

The authors justify their parsimony approach by stating "… we contrast the results of our analyses with those of K15 as a cautionary case study, showing how traditional phylogenetic methods can reveal patterns and suggest explanations that may not be evident to readers, reviewers, or even researchers themselves, from the results of more intricate model-based approaches." But it is well known that parsimony is highly unreliable when there is a lot of homoplasy in the data, particularly likely in DNA sequence data given that for example transitions tend to be considerably more abundant than transversions. I'm not going to get into a parsimony vs. likelihood argument here, and neither do B16, but the authors are clearly attempting to push the argument through the back door, as it were. A parsimony analysis showing something different to a properly formulated model-based analysis is simply not going to be good evidence for error to most of todays' systematists.

As far as support is concerned, the argument that parsimony-based "support values also have the advantage of not being constrained by an upper bound (such as 100%), so the relative strength of strongly supported nodes can be assessed" seems weird. Attempting to argue against probabilistic methods of statistical inference (whether frequentist, likelihood based, or Bayesian) in favor of a method that simply does not allow such comparison (a) is equivalent to arguing against around 130 years of statistics since its development by Pearson and Fisher (to say nothing of the Reverend Bayes), and (b) strikes this reviewer as simply illogical. You might well argue, in the same vein, that a lower bound (0) is also annoying because it doesn't allow comparisons between weakly supported hypotheses. What nonsense!

I think B16 do have a point that some of the mitochondrial DNA data was incorrectly transposed by K15 in the final data matrix. The other nuclear "identical sequences" in allopatric taxa are also troubling, although these were not investigated in as much detail by B16. It would be worthwhile re-running the major analyses of K15 to test what difference careful data checking makes to the conclusions, particularly on network analysis and analyses discussing phylogenetic discordance and

reticulation. However, using the corrected mitochondrial data, B16 have already shown that there's no difference in the overall topology, so it seems likely that this will not change much if some of the suspect nuclear data are corrected.

2) As far as I can see, reticulation was a minor finding of K15, who were mainly interested in producing "the first comprehensive, time-calibrated phylogeny of the group to test hypotheses of diversification rate…" It's possible that, in the K15 abstract, "Despite the large extent of reticulate signal" might mislead some to a conclusion of gene flow between species, but what this really refers to is the pattern found in an analysis of networks (this analysis was not repeated by B16). If anything, the conclusions of actual evolutionary reticulation to explain the phylogenetic discordance observed implies that it was not a central finding in K15. For example: "Nonetheless, consistent with the recent radiation of the group, large effective population sizes and known hybridization between many species, we observed high heterogeneity among sampled fragments of the genome that differ markedly in both topology (Fig. 2, Supplementary Fig. S2, available on Dryad at http://dx.doi.org/10.5061/dryad.44b4j) and rates of evolution (Concaterpillar analysis)." Note, that the invocation of "large effective population size" implicates extensive incomplete lineage sorting, and this is always held to be an alternative to introgression post-speciation. The statement: "The small number of nucleotide differences between the biological species is consistent with interspecific gene flow, which led to a reduction in differentiation across the genome" specifically refers to the melpomene-cydno group. In the melpomene-cydno group, there's an abundance of evidence of hybridization among the species and interspecific transfer of color patterns from other papers, which evidence B16 do not here attempt to argue about (although, due to various catty assides in the paper, as well as previous attacks on studies of gene flow by Brower in Genetica and Proc Roy Soc B, one can infer that they're skeptical about it).

It does seem clear that phylogenetic discordance **is evidence** for evolutionary reticulation, even if it **does not prove it** due to the potential for convergent evolution and incomplete lineage sorting. The B16 argument that the evidence for reticulation goes away when taxa with missing data are removed seems bizarre, given what is known about introgression in the Heliconius melpomene-cydno-pachinus-timareta group in particular.

I'd suggest that if the B16 paper could be toned down, and maybe shortened to a note, and that the authors should try not to take on the whole of model-based likelihood and Bayesian statistics in this paper. It could be a useful critique of the phylogenetic hypothesis supported by K15. The assertion that the use of probability as a measure in statistical inference is somehow wrong because when there is strong support, the probability cannot go to more than 100% should be scrapped, as it contradicts the whole point of statistical reasoning. (The authors are complaining, I think, that such statistical methods are – statistically speaking – "consistent" – i.e. that when there is a lot of data, the correct inference is found by the method. Statistical consistency is what everyone else using statisticistical inference wants. However, I think that the chimeric or transposed data identified by B16 are a potential problem, and it should be tested whether the reticulations demonstrated by network analysis in K15 are strongly reduced when data error are removed. (Not done by Brower & Ordunez). By all means add a parsimony analysis if the authors insist, but they should also try to replicate the major K15 analysis to test for the stability of the hypothesis. Furthermore, I think that the K15 authors should be given right of reply to this revised B16 critique in the journal, since B16 is largely set up as a critique of K15 rather than making any other very general points of interest to Systematics Biology readers.

In the text:

p. 13: The argument that gene flow should create a lot of homoplasy between gene partitions, but little homoplasy within gene partitions is an interesting one, but is I believe incorrect. This might be true if introgression always led to complete replacement of the recipient's sequence by foreign sequence. However, it won't be true if there's a lot of convergent evolution (as is likely for non-coding changes in DNA data for example), or gene flow with no complete replacement, for example if the gene flow is neutral, and if recombination occurs within the partition (as is very likely within insects). As an example, most Eurasians have around 1-2% of SNP variants that came from Neandertals, but they are different variants in different individuals (except in a few notable cases where introgression has been driven by selection).

p. 23: "K15 made a major point of emphasis the provocative claim that traditional systematic approaches are likely to fail to produce a robust, resolved and testable phylogenetic hypothesis for this group of butterflies due to 'conflicting signal' among different gene regions sampled, caused by 'gene flow, hybridization and ILS [incomplete lineage sorting]' (p. 510)." I don't think this is a particularly "provocative" claim any more, and perhaps B16 should stop arguing it's an unreasonable one. With five or more taxa in certain configurations, there's the possibility of an "anomaly zone" where a concatenation of gene trees will strongly support an incorrect branching pattern (Degnan & Rosenberg). For this reason, phylogenetic advances have concentrated in recent years on phylogenetic species tree analysis based on coalescent theory to deal with incomplete lineage sorting (Edwards 2009). But even these methods are still in their infancy, and do not take introgression into account. Notably, they don't deal with the problem of recombination (and see my argument to p. 13 above to show that the problem may be real in these data), and most existing methods have to take short cuts by building (error-prone) gene trees first and then amalgamating these into species trees via a coalescent method. But the need for these methods, however imperfect or ineffective at the moment, cannot be negated by an argument based on a parsimony analysis of concatenated data. Concatenation will cryptically embed incomplete lineage sorting errors into the species tree.

p. 24: "… none of these provides compelling evidence … that the tree from the concatenated dataset is incorrect. K15's point was that the tree from the concatenated dataset is LIABLE to be incorrect, so their finding that in this case it wasn't is important. Also, they noted the failure of the coalescent-based methods to produce very well-supported trees, even though Concaterpillar "rejected concatenation" (p. 25) which is another important result here; some of this could be because, as I've already alluded to, the methodology is still in its infancy; but it's also likely that introgression is also to blame for some of the problems, in the nuclear data at least. Yes, the MSC methods were intended to overcome the problem of incomplete lineage sorting, but not introgression and gene flow. Also, Zhang et al. 2016 show that the whole Z chromosome has a different average tree than the concatenated autosomal tree, showing that there is indeed strong discordance in the data at the genomic level, even disregarding the myriad other evidence of introgression from the Heliconius Genome Consortium 2012, Martin et al. 2013, and Zhang et al. 2016 among other earlier papers.

p. 25: "we and K15 agree that concatenation represents the most plausible approach…" That isn't what I interpret K15 to say. What they're saying is that in this case a BEAST* analysis gave the same tree as a concatenation analysis; they do not say that we should avoid MSC analyses.

p. 26: "K15 repeatedly conflated pattern and process, assuming that incongruence is evidence of reticulate evolution due to gene flow via interspecific hybridization." They didn't. They just said that the known existence of introgression in this group meant that it was one of several possible explanations for the incongruence.

"if sequence manipulation mistakes ….occurred elsewhere in the K15 data or in other datasests used to support iconoclastic evolutionary models endorsed by K15 and others … " K15 makes no such "endorsement," and indeed an endorsement is hardly likely to convince anyone. All they said is that it was a likely explanation. "iconoclastic": What planet did these authors come from? Darwin's finches, sunflowers, sticklebacks, horses, prokaryotes, Drosophila, and even our own species among many others have recently been revealed to show similar effects.

"biologically onerous interspecific gene flow": In Bayesian terms, these authors essentially place a very low, near zero prior on introgression, so that almost any analysis of the data will fail to show a high posterior probability for introgression, whatever the likelihood. But there are hybrids in the wild. These hybrids can backcross. These introgressions are probably rare events on a per individual basis, but rare introgression at a rate of m= 0.01%-0.1% per generation between two species can be shown to be sufficient to cause the observed genomic effects of introgression that have been suggested, and tested extensively in the literature of multiple animal and plant species. So the possibility of introgression is far less "biologically onerous" today than B16 assume. One wonders whether their anger against introgression comes from an outmoded, almost pattern-cladistics point of view, and is based more on idealistic wishful thinking for simpler, earlier days when parsimony was king, than on data or statistical analysis.