Re: trait evolution on networks

3 messages

Brian O'Meara <bomeara@utk.edu>   Wed, Oct 21, 2015 at 11:20 PM
To: Cécile Ané <cecile.ane@wisc.edu>
Cc: Paul BASTIDE <paul.bastide@agroparistech.fr>, Alison D Scott <alisondawnscott@gmail.com>, Claudia Solis
Lemus <claudia@stat.wisc.edu>, Mohammad Khabbazian <khabbazian@wisc.edu>, Tony Jhwueng
<tony.jhwueng@gmail.com>

Ah, that makes sense. Inheriting from two parents will tend to decrease variance. Thank you also for picking
up on the other error in the code and for pointing out the Pickrell & Pritchard paper. Would you and Paul mind
if we thank you in the acknowledgements [and should we include someone else -- or just say the X lab
group]? [some journals now require that people acknowledged agree -- I think this is to prevent "We thank
Drs. X and Y, whose terrible work inspired our much better approach"].

The first version of the paper was rejected; after working to improve it, in August we submitted the version you
read. We're still waiting to hear back, but I'm sure there will be ample opportunity to fix this before the paper is
published (and of course, even if they say publish it immediately (ha) we'll fix these errors and rerun
everything before publishing).

Thank you all for your close reading of this. It's the sort of thing one hopes peer reviewers do, but too often
errors get missed by authors and reviewers.

Best,
Brian

PS: Cool paper on networks. I like that you do not assume that hybrid edges have length zero -- it's something
we also do, but it does make it more complex. It's far better than assuming full sampling of all species that
were involved in hybridization at some point in the past. It's too bad (but understandable) that terminal branch
lengths can't be estimated if you do one sample per species; perhaps one future extension could be to fit a
model between gene tree branch lengths and internal estimated branch lengths in coalescent units on the
species tree and then use this to predict terminal branch lengths given the gene tree edges for them (perhaps,
in both internal and terminal, filtering for gene trees that match the local species tree topology for the region
around the edge). One issue we've had with our paper on approximate likelihood for phylogeography
(http://biorxiv.org/content/early/2015/08/24/025353), which also uses gene tree topology only, is reviewers
complaining that we're throwing away information. This is true, of course, but in our case we argue that it
dramatically simplifies the problem and that branch lengths on individual gene trees are likely to be very noisy,
so they probably wouldn't provide much useful signal. Perhaps this argument may help in your case, too.

_____

Brian O'Meara
Associate Professor
Dept. of Ecology & Evolutionary Biology
U. of Tennessee, Knoxville
http://www.brianomeara.info

Postdoc collaborators wanted: http://nimbios.org/postdocs/
Calendar: http://www.brianomeara.info/calendars/omeara

On Tue, Oct 20, 2015 at 11:06 PM, Cécile Ané <cecile.ane@wisc.edu> wrote:
  Hi Brian,

  I hope things are going well for you these days!
  We read your paper with Dwueng-Chwuan Jhwueng as a group (http://biorxiv.org/content/
  early/2015/08/05/023986), and it seems to us that there might be some errors in the covariance formula for
  the diagonal terms. In your small example page 7-8, our derivations for the variance of R gave this:
  $\sigma^2*(t\_1+t\_2)*(m^2 + (1-m)^2) + sigma\_H^2 * t\_3 + \nu\_H$.
  The difference with your formula is in the $m^2 + (1-m)^2$ factor. Paul Bastide is visiting UW-Madison this
  Fall and lead our discussion. He checked in your BMhyb package to see if that was just a typo in your

preprint, and if you had the same formula we found in your implementation. But he found that the diagonal terms matched the formula in your preprint, and that instead some off-diagonal terms did not (there were too large compared to what was expected in the formula). We didn't investigate things further ---sorry I don't have code to be more specific about what Paul found. But I just wanted to let you know.

We might be wrong about the formulas, but what we derived actually agrees with the trait evolution model already described by Pickrell & Pritchard (2012) for networks: Inference of Population Splits and Mixtures from Genome-wide Allele Frequency Data (PLOS genetics 8(11):e1002967). They focus on frequency data, so we just need to replace their ancestral x(1-x) by a sigma^2 parameter. But they provide a general formula for the covariance matrix, given any network. It's in their Supplementary Information text, page 3 mostly, equations (12) and (13). Even though their goal is not to study the evolution of a particular trait (instead their goal is to reconstruct the network from many many traits), it's the first paper that I know of where such a trait model is presented on a network. Very cool!

If it's not too late for your paper with Dwueng-Chwuan, perhaps it's worth a check for these typos and their implementation. Perhaps you will find that the estimated parameters get closer to the true values in your simulations! In any case, please correct us if we are wrong. By the way, my student Claudia has a cool method (I think, but I know that I am biased) to estimate a network from multi-locus sequence data: http://arxiv.org/abs/1509.06075 (and Julia software https://github.com/crsl4/PhyloNetworks). These network models are exciting, I think.

Cheers,
Cecile.

--
Cécile Ané
Departments of Statistics and of Botany
University of Wisconsin - Madison
www.stat.wisc.edu/~ane/

CALS statistical consulting lab:
www.cals.wisc.edu/calslab/stat_consulting.php

Brian O'Meara <bomeara@utk.edu>   Wed, Oct 21, 2015 at 11:21 PM
To: Tony Jhwueng <tony.jhwueng@gmail.com>

Hi, Tony. Any word on our paper from Syst Bio? It could be worth an email to the AE at this point.

Best,
Brian

_____

Brian O'Meara
Associate Professor
Dept. of Ecology & Evolutionary Biology
U. of Tennessee, Knoxville
http://www.brianomeara.info

Postdoc collaborators wanted: http://nimbios.org/postdocs/
Calendar: http://www.brianomeara.info/calendars/omeara

[Quoted text hidden]

Cécile Ané <cecile.ane@wisc.edu>   Wed, Oct 28, 2015 at 4:06 AM
To: Brian O'Meara <bomeara@utk.edu>
Cc: Paul BASTIDE <paul.bastide@agroparistech.fr>, Alison D Scott <alisondawnscott@gmail.com>, Claudia Solis Lemus <claudia@stat.wisc.edu>, Mohammad Khabbazian <khabbazian@wisc.edu>, Tony Jhwueng <tony.jhwueng@gmail.com>

Hi Brian,
Sorry for my delayed response! Things pile up too fast. Of course you are welcome to acknowledge us. I hope your PHRAPL paper will be well received this time!

Cecile.
[Quoted text hidden]

--
Cecile Ane
Departments of Statistics and of Botany
University of Wisconsin – Madison
www.stat.wisc.edu/~ane/

CALS statistical consulting lab:
www.cals.wisc.edu/calslab/stat_consulting.php