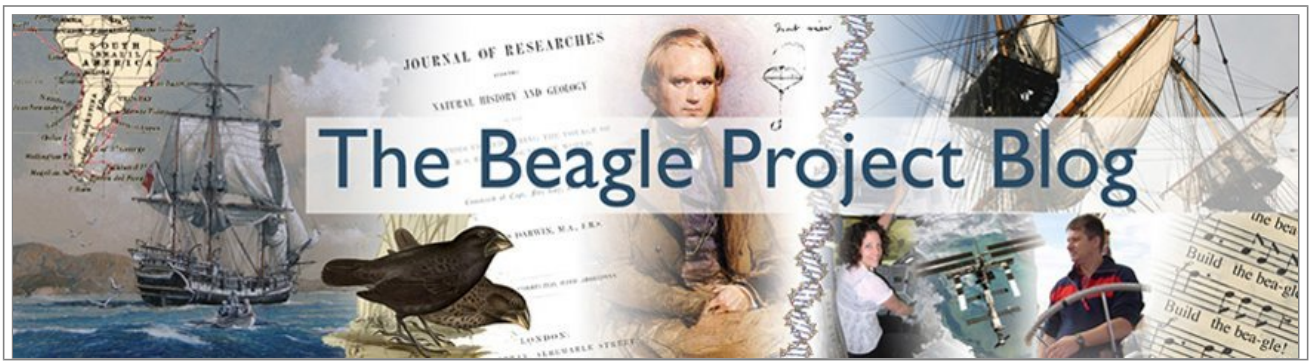


HELP US REBUILD THE BEAGLE!

Donate in $US:

Donate in £GB:

MINI-BLOG

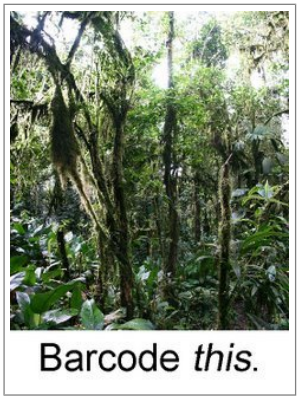
As T. Ryan Gregory writes, "There already are qualified people who constantly challenge evolutionary explanations. They are known as evolutionary biologists." *-kj*

Get ready for the Perseids on August 12th in a pre-dawn sky near you. *-kj*

17 FEBRUARY 2008

Would that which we call a rose, by a DNA barcode, smell as sweet?

For my second post on peer-reviewed research (I suppose I'll have to stop introducing it like that after three or four), I've chosen a paper that gets right to the beating heart of my own tiny little corner of science geek-dom: DNA barcoding in plants.


Barcode *this*.

To make things more interesting, the science press worked themselves into a premature and, as I will argue here, seriously specious frenzy last week when they collectively oohed and ahhed about the paper in terms that were well, let's just say, um, how do I put this delicately ...flat wrong.

So first, as is right (though apparently not at all customary) to do before trumpeting a paper's "conclusions" far and wide, why don't we have a look at the paper itself?

In the article (open access in PNAS), Renaud Lahaye and colleagues of the University of Johannesburg, with co-authors from Lankester Botanical Garden in Costa Rica and Imperial College London, and senior (and corresponding) author Vincent Savolainen at Royal Botanic Gardens, Kew, report the results of their significant new evaluation of eight proposed DNA barcodes in plants.

A DNA barcode is a short snippet (less than 600 base pairs) of a organism's native genome (billions of base pairs) that should contain enough unique information to accurately identify an unknown specimen (for example, a piece of mystery meat in a Japanese ~~whale~~ fish market) to the species level. This is to be done by comparing the sequence of the query snippet against a comprehensive database of sequences from specimens that have been identified and vouchered by expert taxonomists.

Importantly, a DNA barcode is not just any unique snippet of an organism's DNA, but the *same* snippet in all organisms, slight variations within which provide the sought-after species-specific signal. In other words, all DNA barcodes are ultimately descended from the same gene that was present in the common ancestor of all living organisms. Uh oh, creationists aren't going to like *that*, are they? They're especially not going to like the fact that *it works*.

Well, to be more accurate, it works for *animals*. See, the first big question that has had to be addressed to implement DNA barcoding as a common procedure for identificaiton is: which

Tall Ship stuff on UK TV (5.8.08): why did the Mary Rose sink? Channel 5 8pm. pmc

Olivia Judson shoots and scores with yet another brilliant piece on evolution in the New York Times. *-kj*

What do the Galapagos islands and my home state of Colorado and have in common? Find out in "Where Research and Tourism Collide" by Michelle Nijhuis in the New York Times. *-kj*

A big Beagle Project thank-you to Peter Etnoyer of Deep Sea News and Rick MacPherson of Malaria, Bedbugs, Sea Lice and Sunsets for responding on incredibly short notice to a barrage of interview questions for a Spanish newspaper article about El Proyecto HMS Beagle. As soon as it's out you can bet we'll be posting it here. *-kj*

Jason-2 ocean topography satellite launches in California. *-kj*

Beyond Cladistics: A festschrift for Professor Chris Humprhies, 1-3 October 2008 at the Linnean Society of London (website, programme pdf) *-kj*

If you liked Bio-Rad's *The PCR Song*, you're gonna love Eppendorf's *epMotion* *-kj*

On the main website of the Natural History Museum in London: 1) Student summit-Darwin and contemporary science, 2) Darwin and Wallace webcast at 12:30 GMT+1, 3) Exhibition: "Happy birthday natural selection" *-kj*

YOUR BEAGLE BLOGGERS:

Karen James

Peter Mc

snippet? Finding a good DNA barcode is more difficult than it sounds. It has to have undergone just the right amount of mutation during its long evolutionary journey such that the differences between species (interspecific variation) are not swamped out by the differences within species (intraspecific variation). Unfortunately, finding a gene that harbours this sweet spot of variation, called the "barcoding gap", is like finding a needle in a haystack.

It was with great excitement, then, that in 2003 a group of Canadian scientists reported that a gene called cytochrome oxidase 1 (CO1) seemed to have just the right amount of variation. They showed that just about every species of animal has a unique *CO1* sequence, and that the "barcoding gap" was nice and big (though serious doubt was later cast on this claim when Chris Meyer and Gustav Pauley showed that in three groups of marine gastropods the "barcoding gap" was an artifact of insufficient sampling). This meant that by comparing the sequence of an unknown animal's *CO1* gene back to a reference database of *CO1* sequences built up from all known animals (which, as Meyer and Pauley argued, had better be extensively sampled and taxonomically sound), we would have an excellent chance of working out the species identity of the unknown animal.

Four out of five cowries agree: *the barcoding gap ain't all it's cracked up to be.*

Well that's just dandy for all those zoocentrics out there, who have already databased barcodes from hundreds of thousands of vouchered specimens, but what about plants? I mean, wouldn't it be great to be able to validate the identity of herbal extracts or rapidly survey the diversity of dormant seeds in the soil in a proposed conservation area? Well, unfortunately, *CO1*, though present in plants, is not variable enough in land plants to use in species identification (though it does seem to work for some marine algae). Land plants also pose other problems, like hybridisation, duplicated genomes and asexual reproduction, which may mean it's not even *possible* to find an ideal DNA barcode in plants. In other words, the search is on.

Though there's a *Science* News Focus on this quest for ~~the holy grail?~~ a universal land plant barcode, it's subscriber-only (boo), so here's a quick list of the plant barcode candidates floated by various teams of investigators over the last three years:

- *trnH-psbA* (Kress *et al*, 2005)
- *rbcL* and something else (Newmaster *et al*, 2006 - pdf)
- *rbcL* and *trnH-psbA* (Kress & Erickson, 2007)
- 1) *rpoc1*, *rpoB*, and *matK* or 2) *rpoc1*, *matK*, and *trnH-psbA,* with an honourable mention going out to *accD*, *ndhJ*, and *YCF5* (Chase *et al*, 2007 and the Kew DNA barcoding website)
- *trnL* intron (Taberlet *et al*, 2007)
- *matK*, *trnH-psbA* and *atpF-atpH* (proposed based on unpublished data by Ki-Joong Kim at the very lively Plant Working Group meeting at the 2nd International Barcode conference in Taiwan in September; this combination was in the lead at the end of the session and a vague agreement was made to follow it up)

As you can plainly see, plant barcoding is, as a discipline, one big ol' entangled bank. So what's a wannabe botanical barcoder to do?

Lahaye *et al* seem to have opted for the bigger is better approach: they tested all of the above (with the exceptions of the *trnL* intron and *atpF-atpH*) on a more challenging set of plant specimens than had yet been tested (previous barcode trials have necessarily had to examine a

Elke Watts

BEAGLE PROJECT AUDIO & VIDEO

Full screen view

BEAGLE PROJECT IN THE NEWS

Note: for radio interviews and other audio, use the Beagle Project Audio Player above

2008
LabLit
Nature Network News
Science Creative Quarterly (Peter brings Darwin back from the dead to comment on the Beagle Project; artwork by Diana Sudyka)
A Blog Around the Clock

2007
Science
Yachting Monthly
Times Online

2006
BBC News

More Beagle Project media links here including local and international coverage.

very broad taxonomic range with minimal in-depth sampling within species, and have not attempted to discriminate specimens from particularly species-rich areas that are likely to cause the most difficulty). As the authors put it, “the critical test of evaluating the applicability of DNA barcoding for biodiversity inventories in species-rich geographic areas has been lacking.”

Lahaye *et al* test the eight barcodes from specimens from not one but two of these "species-rich" areas: Costa Rica, where they focus on orchids, and Kruger National Park in southern Africa, where they focus on trees and shrubs.

On first glance, it seems that they have collected and tested a truly eye-popping number of specimens (1,667!) but as you read the paper more carefully you see that 1,495 of these were orchid *matK* sequences "collected" from GenBank (which, it must be said, is notorious for its bad taxonomy), and only the remaining 172 specimens (101 southern African trees and shrubs and 71 Costa Rican orchids) were tested against all eight candidate barcodes. Nevertheless, it's still a (slightly) more comprehensive effort than has previously been carried out, and the increased "sampling" provided by the GenBank accessions does seem to enable a robust examination of the "barcoding gap" for *matK*.

So, here is a rapid-fire summary of the main results:

1. With the exception of *ndhJ* an *ycf5* in orchids, PCR amplification reactions were successful. This is not just a methodological afterthought, but rather an important result considering the fact that finding good, universal primer pairs (especially for *matK*) has been a real problem for the plant barcoding community.

2. Inter- and intraspecific genetic divergences were calculated, and the size of the "barcoding gap" assessed for each candidate barcode (I was very pleased to see that they used the Meyer and Pauley metrics for this). Both *matK* and *trnH-psbA* performed fairly well here (i.e. of the eight barcode candidates they had the highest inter- and lowest intraspecific divergences), though, as Meyer and Pauley would have predicted, no large barcoding gap was found. That said, analysis of the large matrix of (mostly GenBank) orchid *matK* sequences revealed a pretty darn good barcoding gap if you ask me:

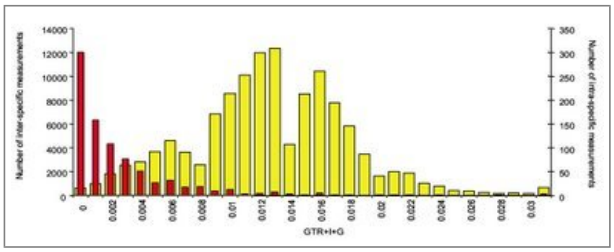


Figure 1I from Lahaye *et al* shows the "barcoding gap" between interspecific divergence (yellow) and intraspecific distances (red) in the *matK* gene among Costa Rican orchids.

3. *matK* sequences were able to detect "cryptic species" (real species boundaries masked by physical similarity) in the orchid data set, certainly a good sign, since that is one of the eventual utilities of a DNA barcoding system. For example, one of their four samples of *Lycaste tricolor* did not cluster with the other three as expected and thus it *may* be another, separate species. To find out if it actually *is* another species, some *real* taxonomy will need to be done.

ABOUT THE HMS BEAGLE PROJECT

(en Español: El Proyecto HMS Beagle)

We aim to celebrate Charles Darwin's 200th birthday by building a sailing replica of HMS *Beagle* and then retracing the 1831-1836 Voyage of the *Beagle* with an international crew of researchers, aspiring scientists and science communicators. The new *Beagle* will symbolise both the physical and intellectual adventure of science; she will be equipped with laboratories, 21st century science equipment and satellite communications, she will host cutting-edge science projects of international relevance while serving as vehicle for improving wider public engagement with and understanding of science. The eye-catching image of a traditionally rigged sailing ship will also embody the drive for a sustainable, fossil fuel free future. If you support our vision, please donate to our build fund:

Donate in $US:

Donate in £GB:

The HMS Beagle Trust is a not for profit UK limited company no. 6025763, charitable status pending.

MORE ABOUT THE BEAGLE PROJECT:

Aug '07 Newsletter:
3.5Mb/332Kb
The Beagle Project home
The Beagle Project shop
Replica plans

matK sequence data pointed to a potential "cryptic species" hiding amongst four samples of *Lycaste tricolor*.

So, in conclusion, Lahaye *et al* argue for the adoption of the *matK* gene as the universal DNA barcode for land plants (with an option for use of *trnH-psbA* as an alternative or complement to *matK*). *matK* had been previously tut-tutted because it was difficult to amplify from many different groups of plants, but this seems to have been overcome here by a particular set of primer pairs (390F and 1326R from Cuenoud *et al*., 2002) which amplified with "100% success".

However, and this I think is one of the key take-home messages here, Lahaye *et al* found that the power of *matK* and/or *trnH-psbA* to correctly identify species was only approximately 90% and that therefore "we may need to accept that no more than ~90% of species will be identified with universal plastid barcodes and that those difficult lineages will need 'case-by-case' analyses, using, for example, nuclear population genetic markers and taking advantage of recent developments in DNA sequencing technology."

Hmm, very provocative. *smiles knowingly*

To wrap things up here, I'd like to end where I began, with the premature media hype that preceded the publication of this paper. Various sources claimed that in this paper, a DNA barcode for plants has been "mapped", "revealed", "finally revealed", "found" "identified", "determined", "decided", or even just simply "is". Based on what you've read here, do *any* of these sound even remotely accurate?

Don't get me wrong, this paper is an important contribution to plant DNA barcoding, which is why I have chosen to blog about it in some detail here, but it is, in essence, just another proposal in a long string of proposals. To the authors' credit, they did *not* claim to have made The Final Decision on the identity of the plant DNA barcode(s), but the press sure did.

So, to get the bad taste of sensationalist hyperbole out of our mouths, I thought I'd leave you with some nice minty-fresh alternative headlines. How about:

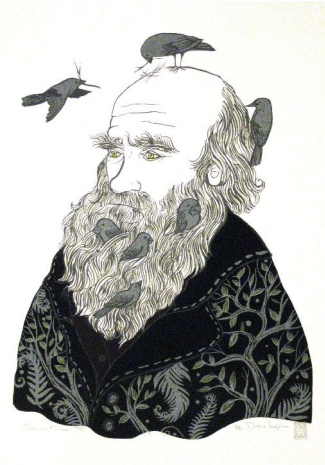
"Candidate DNA barcodes for plants tested in largest study thus-far."

Or, if it must be short and sweet:

"DNA barcodes for plants tested" or "Plant DNA barcode proposed".

Now, how hard was that?

Voyage of the *new* Beagle
Science aboard
Beagle in the media
About us
Contact us



BIG BROTHER

TEN MOST RECENT POSTS BY BEAGLE BLOGROLLERS (SEE COMPLETE LIST BY CLICKING "SHOW ALL" AT BOTTOM)

The Other 95%
SOS - The Most Dangerous Species
1 hour ago

Laelaps
Photo of the Day #340: Lion
1 hour ago

Pharyngula
Funnels and Tornadoes and Lightning, Oh my!
1 hour ago

A Blog Around The Clock
ClockQuotes
3 hours ago

Lahaye, R., van der Bank, M., Bogarin, D., Warner, J., Pupulin, F., Gigot, G., Maurin, O., Duthoit, S., Barraclough, T.G., Savolainen, V. (2008). DNA barcoding the floras of biodiversity hotspots. *Proceedings of the National Academy of Sciences DOI: 10.1073/pnas.0709936105*

Hebert, P.D., Cywinska, A., Ball, S.L., deWaard, J.R. (2003). Biological identifications through DNA barcodes. Proceedings of the Royal Society B: Biological Sciences, 270(1512), 313-321. DOI: 10.1098/rspb.2002.2218

Meyer, C.P., Paulay, G. (2005). DNA Barcoding: Error Rates Based on Comprehensive Sampling. PLoS Biology, 3(12), e422. DOI: 10.1371/journal.pbio.0030422

POSTED BY KAREN JAMES AT 23:19

LABELS: DNA BARCODING, SCIENCE BLOGGING, SPLENETIC ANNOYANCE

5 COMMENTS:

Sid Leavitt said...

Wow again. Keep 'em coming. And, by the way, thank you again.

18 FEBRUARY 2008 14:22

Richard Carter, FCD said...

Another very interesting post, thanks. I had no idea we don't yet have a reliable barcoding technique for plants.

18 FEBRUARY 2008 20:43

Becca said...

Great post!
I'm not terribly familiar with the current literature on this topic- but I remember back in the day Woese suggesting using ribiosomal genes as true universal identifiers. How does the barcode concept differ from that?

19 FEBRUARY 2008 15:30

nunatak said...

Hi becca, thanks for your comment (and for enthusing over on the Clock). This isn't really any different from Woese's suggestion. DNA barcoding is really just a new name pasted onto that idea ...but it's not the idea here that's so important but the will and large scale cooperation needed to make it a reality. It's an organised, standardised, international approach to actually implement Woese's dream. There's more on Woese and the history of barcoding here: http://en.wikipedia.org/wiki/DNA_barcoding#Origin

20 FEBRUARY 2008 18:11

Felicia Gilljam said...

Nice post. Media is silly like that. And unfortunately, scientists seem to constantly cater to that. If you want attention, you better write a sensationalist press release.

I remember reading about a palm tree on BBC News a month ago or so where they kept exclaiming it was "self-destructing" because it lived for a hundred years and then suddenly flowered and died. As if dying after a strenuous bout of reproduction is in any way exceptional

among plants, or for that matter other organisms. But how else to get media interested in a tree?

21 FEBRUARY 2008 16:23

Post a Comment

LINKS TO THIS POST

Create a Link

Newer Post Home Older Post

Subscribe to: Post Comments (Atom)

Charles Darwin's Beagle Diary
13th August 1833
6 hours ago

Tangled Up in Blue Guy
Carnival of the Liberals LXXI
7 hours ago

Greg Laden's Blog
Carnivalia!!!
9 hours ago

The Dispersal of Darwin
Darwin & The Dodo BW
15 hours ago

Cephalopodcast
Goblin Shark Bites Scuba Diver (Mitsukurina owstoni)
15 hours ago

The Oyster's Garter
Why nearshore oil drilling is bad: a handy guide
16 hours ago

Show All

Looking for an older post? Visit the archives from our old blog address (warning: lots of the pretty pictures are gone, lost to the void that is cyberspace, but we like to think it's still riveting reading).

MERIT BADGES

What is this?

WORTHY CAUSES

BEYOND CLADISTICS
A FESTSCHRIFT FOR CHRIS HUMPHRIES

1st - 3rd October 2008

BLOG ARCHIVE

▼ 2008 (202)

▼ August (13)

Glendon Mellow (aka the