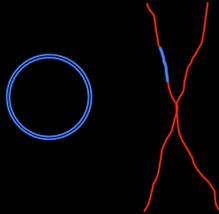


Problems with numts (nuclear copies of mtDNA)

Kathy Hill



what are numts?

- numts are copies of mitochondrial DNA found in the nuclear DNA
- also known as pseudogenes, homologs or paralogs
- how does the mtDNA get into the nuclear DNA?

size and number of numts

- within a taxon, mtDNA fragments can be transferred multiple times, in multiple different sizes
 - usually thought to be small (<1000bp), however a lot of evidence now for large many kb numts
- once inside the chromosomes numts are subject to being copied again
 - for example 7.9kb of mtDNA were transferred to the *Felis sylvestris catus* chromosome D2, and then tandemly repeated 38-76 times
- numts appear to be transcriptionally inactive
(nuclear genetic code is different to the mtDNA genetic code)
- numts appear to be much more common in some organisms than others

divergence of numts

- divergence between numt copies and real mtDNA copies
- inside the nuclear DNA, the numt
 - should change much slower than the real mtDNA copy
 - Should have changes that are random hits, not 3rd base weighted and can include indels (that are not in sets of three)

amplification of numts

- if large enough mtDNA fragments are transferred, then the primer sites for gene amplification will be copied, and numts can be amplified
- numts can be preferentially amplified using universal primers
for example the orangutan
- amplification of numts seems to be easier when DNA is degraded, such as when amplifying from dried specimens
 - why?
possibly because the mtDNA are less well-protected than the nuclear DNA

how do you know if you've got numts?

- double bands in PCR or double peaks in mtDNA sequences
(but need to rule out mtDNA heteroplasmy)
- if an mtDNA gene is heteroplasmic
 - you should be able to find individuals with single copies of both alleles
 - both copies must be able to be transcribed (no stop codons)
 - copies should be closely related

why are numts a problem?

-mixed copies of numt DNA and real mtDNA in the same dataset may infer incorrect relationships when building phylogenies

-for distantly related taxa, the numt may not be a particularly big problem for phylogenetic placement, because the numt copy is probably younger than the division with the next relative.

-however the numt sequence can reconstruct an incorrect (shorter) branch length because less evolution is being inferred

-in phylogeographic (intraspecific) studies numts are especially a problem, because the numt sequences may have diverged before any of the within species splits

accidentally published numts

-numts have been accidentally published as real sequences in several cases that have been discovered, and many more such cases may be discovered in the future as genome studies find numts when the entire genome is sequenced

-for example, human numts have been published as dinosaur DNA and ancient monkey DNA

when are numts useful?

-numts represent ancestral mtDNA, especially in taxa where nuclear genes evolve slowly

-can be useful for rooting phylogeographic studies

-can be used to infer ancestral states

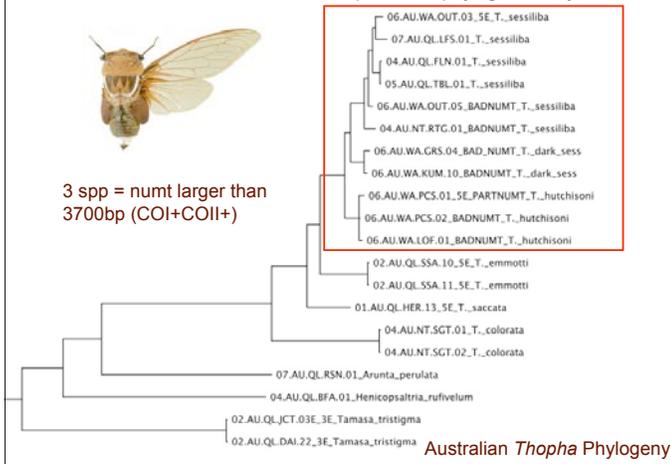
numts in Cicadidae

-in the Simon Lab, numts have been amplified in ?<50 cicadas so far:

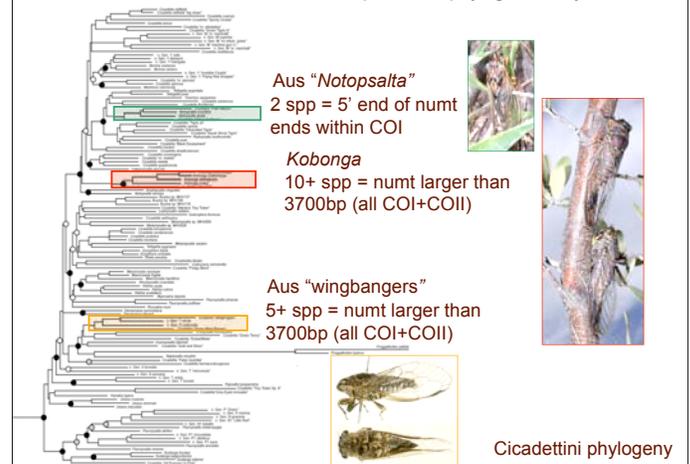
Cicadettinae	Cicadettini	
		<i>Maoricicada</i> (1 sp)
		Australian "Notopsalta" (4+spp)
		Australian wingbangers (2+gen, 5+spp)
		<i>Cicadetta</i> (2 spp)
		N. gen. <i>lineola</i> (1 sp)
		<i>Kobonga</i> (10+ spp)
		<i>Kikihia</i> (1 control lane???)
		"False Quintilia" (1 sp)
		<i>Pauropsalta</i> (a few spp)
		<i>Rhodopsalta</i> (1 sp)
		4 genera, 4 sp
	"Pictilini"	<i>Magiccada</i> (1 sp)
	Taphurina	<i>Burbunga</i> (1 sp)
	Various Aus Tribes	<i>Jassopsaltia</i> (1 sp)
		<i>Talcopsaltia</i> (1 sp)
		<i>Thopha</i> (3 spp)
		<i>Tibicen</i> (9 sp)
	Cryptotympanini	<i>Chremistica</i> (2 sp)
	Psithyristriini	<i>Psithyristria</i> (2 sp)

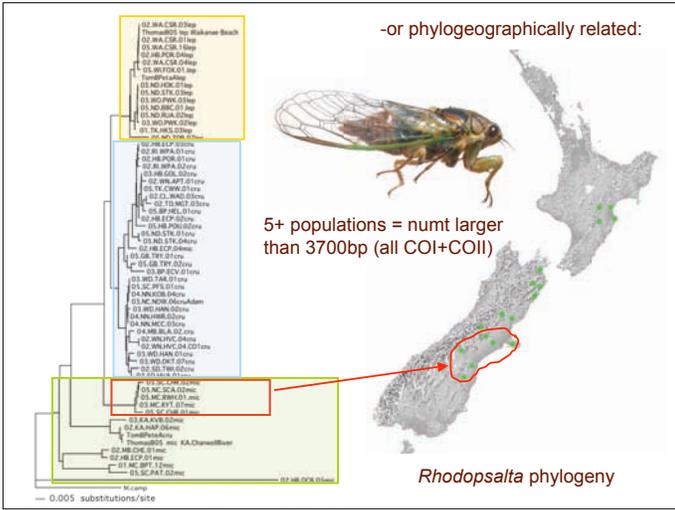


-often individuals where numts are amplified are phylogenetically related:



-often individuals where numts are amplified are phylogenetically related:

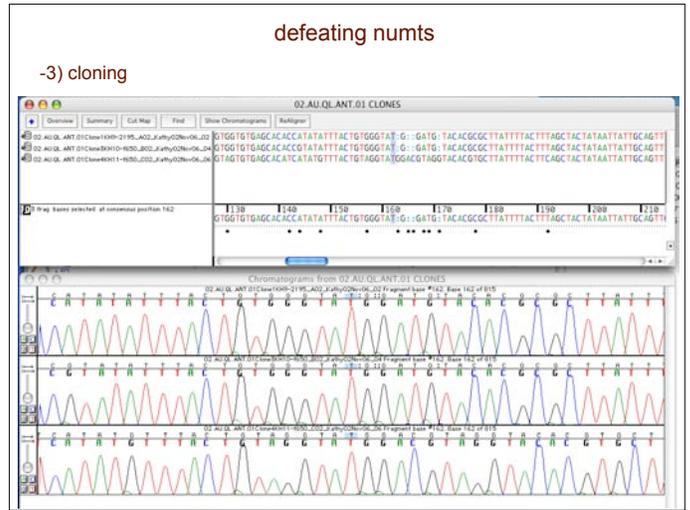
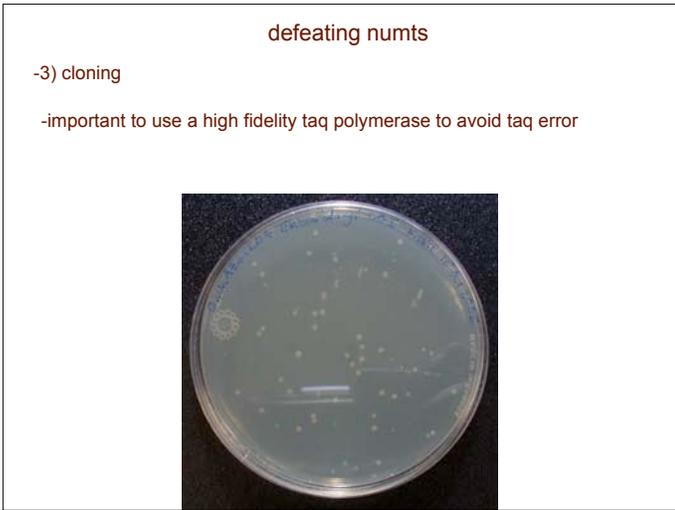
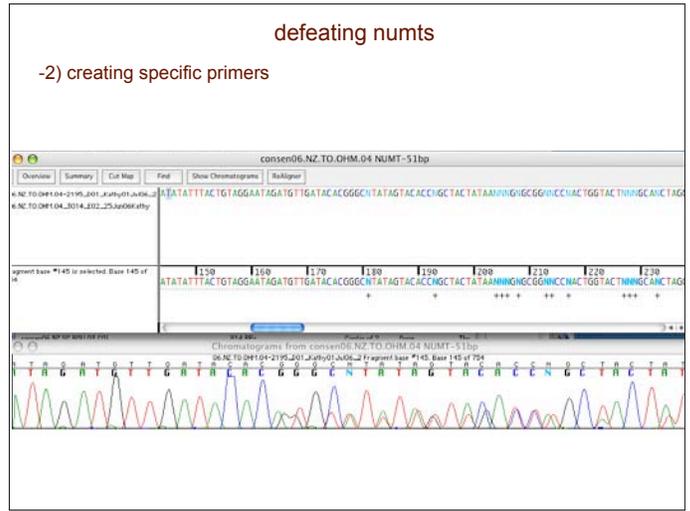
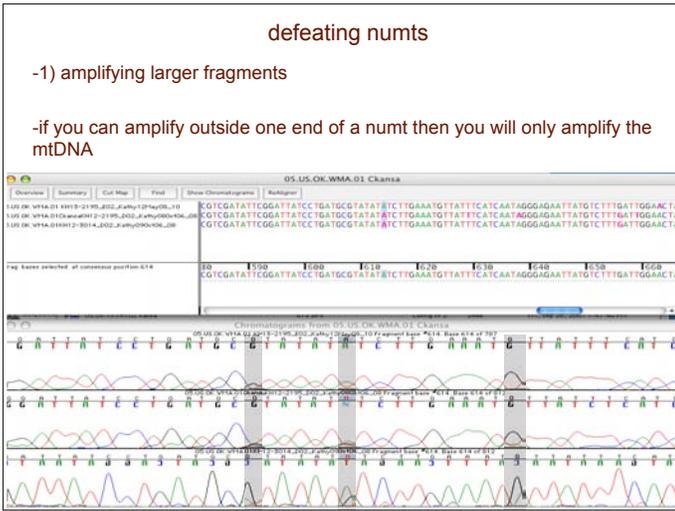




defeating numts

-once you have determined that you have numts (usually due to seeing double peaks in sequences), there are several ways to go about figuring out the correct mtDNA copy

- 1) amplifying larger fragments
- 2) creating mtDNA specific primers
- 3) cloning
- 4) purifying mtDNA



defeating numts

-4) purifying mtDNA

- CsCl more pure but difficult and toxic
- Alkaline lysis - semipure mtDNA but low success rate
 - homogenize tissue in sucrose/glucose buffer
 - centrifuge gently and remove supernatant containing cells
 - centrifuge rapidly to pellet cells
 - add buffers and alkaline lysis solutions
- use a kit or phenol chloroform to extract DNA
- dilute resulting DNA to the lowest amount that will still amplify for PCR

defeating numts

-4) purifying mtDNA

Promega Wizard SV minipreps kit + dilution

used two whole cicada legs and extracted

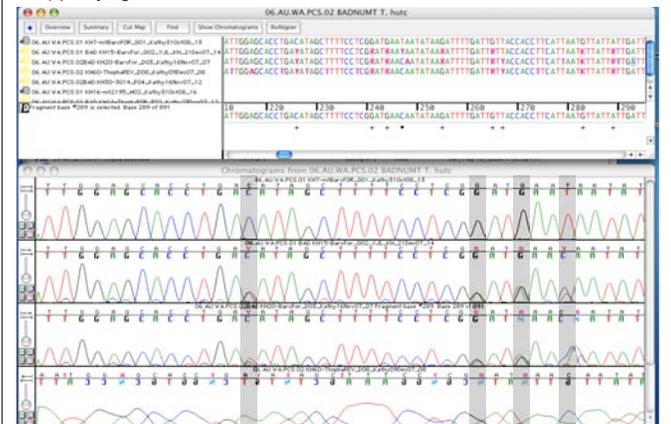
- 2 *Thopha* - successful!
- 1 *Talcopsaltria* - unsuccessful
- 2 *Kobonga* - unsuccessful
- 3 *Tibicen* - unsuccessful

-eggs are the best source of pure mtDNA

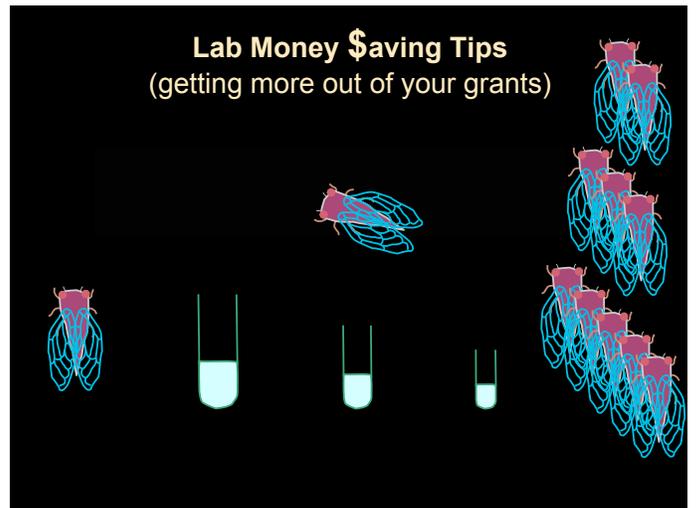
- important to preserve whole bodied samples of all taxa

defeating numts

-4) purifying mtDNA

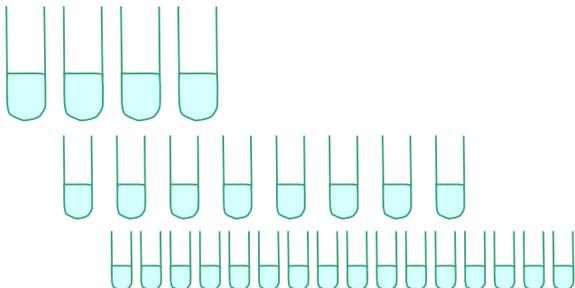


Lab Money \$aving Tips (getting more out of your grants)



saving money in the lab usually comes from one thing:

-use less/do smaller reactions!



PCR:

try to amplify only what you need to visualise on a gel and then cycle sequence

PCR amplifications can work well with as little as 6.25 μ L total volume (this is 1/4 25 μ L PCR)

x-taq kit = \$160

- = 385 25.0 μ L reactions (42c per)
- = 770 12.5 μ L reactions (21c per)
- = 1540 6.25 μ L reactions (10c per)

so if your amplifications are working well try to cut them down

smaller total volumes also means you use less template DNA

gel visualisation

put your EtBr/Sybr in with your sample instead of in the gel when it is cast

1 tube of Sybr Safe = \$~30

2.5 μ L Sybr Safe in 30mL gel

= \$~0.19 for ~24 samples

500 μ L loading dye + 1.5 μ L Sybr

then use 1 μ L per sample

= \$~0.0054 per gel of 24 samples

can also re-use gels with this method, simply run off the old samples and use gel again

(1 30mL gel itself is \$~0.50 for agarose)

PCR clean up

using clean up kits (e.g., Nucleospin Extract II):

-reuse the columns

-after eluting the sample for the final time,

run water (at room temperature) through the column twice

(hint: NT3 Nucleospin buffer is just 76% EtOH!!!)

1 box Nucleospin Extract II kit 250 columns = \$339

- one use per column = \$1.36

- two uses per column = \$0.68

- three uses per column = \$0.45

PCR clean up

using ExoSAP-IT:

-1 tube of 1mL ExoSAP-IT = \$334

= \$0.33 per μ L

-dilute 1:5 ExoSAP-IT:water, then add

1 μ L diluted ExoSAP to 2.5 μ L PCR sample

1 μ L ExoSAP-IT mix = \$~0.07

I usually clean up 2.5 or 5 μ L of the PCR sample

(1 or 2 μ L ExoSAP mixture per PCR)

cycle sequencing

like PCR, try to use a smaller reaction volume

Big Dye = \$~1.00 per μ L

I do a 5 μ L cycle sequencing reaction per sample

1 μ L Big Dye (\$~1.00)

0.5 μ L Big Dye buffer

0.75 μ L primer @ 1 μ M

water+template to 5 μ L

small volume cycle sequencing also allows you to maximise your PCR

- if you have poor amplification it is easier to have enough template if you use a smaller cycle sequencing volume

- (I sometimes do 2.5 μ L reactions with 1/2 μ L Big Dye and so far they work)