

Practicals Biology – Virginie Courtier-Orgogozo 2019

I. Red and green aphids

Pea aphids can be green or red. Clones of green aphids (named Gxxx) and red aphids (named Rxx) are maintained via asexual reproduction in the laboratory.

Pea aphids carry several endosymbiont species, including *Buchnera aphidicola*, which is transmitted maternally.

a. Present two possible trajectories for maternally transmitted alleles in pea aphids.

Crosses were performed to understand the genetic basis of the colour difference. Each line of the table represents the result of one cross.

Parental Clone 1	Parental Clone 2	Total # Progeny Clones	# Red	# Green
R 5A	G 8101	60	35	25
R 5A	G 2a	20	12	8
R 82b	G 2a	56	34	22
R 82b	G 921	12	9	3
R 82b	G 2a	8	4	4
G 721	G 921	10	0	10
G A2C	G 921	3	0	3
R 5A	R 5A	9	9	0

b. If you assume that there is a single locus, is the red allele recessive or dominant? What is the genotype of clones R5A, G921 and GA2C?

The green and red colours are due to the presence of green and red carotenoids. Carotenoids are produced by plants, fungi and bacteria whereas animals are unable to produce carotenoids de novo. Accordingly, carotenoid biosynthetic enzymes genes are found in plants, fungi and bacterial, but usually not in animal genomes.

Let's see if the pea aphid genome contains such a carotenoid synthase cyclase gene!

c. Retrieve a carotenoid synthase cyclase gene:

Go to NCBI Protein <https://www.ncbi.nlm.nih.gov/protein/>

Search for "carotenoid synthase cyclase"

"lycopene cyclase/phytoene synthase" is the type of gene we are looking for, let's copy this protein sequence: UniProtKB/Swiss-Prot: P37295.1.

d. Search for possible carotenoid synthase genes in pea aphids:

Go to NCBI BLAST <https://blast.ncbi.nlm.nih.gov/Blast.cgi>

Select the correct type of BLAST. BLASTP is only for annotated proteins.

Select the pea aphid sequence data: "Acyrtosiphon pisum (taxid:7029)"

Run BLAST.

e. For comparison, search for possible carotenoid synthase genes in *Drosophila melanogaster* or *Anopheles gambiae* on two other browser tabs.

f. Compare the results of the various BLAST searches.

Let's examine the hit sequences XM_001943135.5 and XM_016801018.2. On the right of the BLAST search results, click on page click on "[Gene](#) - associated gene details"

You can see that the sequence is found on Chromosome A1 – NC_042494.1, so it is apparently not a contamination.

g. What is the difference between hit sequences XM_001943135.5 and XM_016801018.2?

Carotenoid biosynthesis involves two types of enzymes: phytoene synthase/carotenoid cyclases and carotenoid desaturases. Table 1 and Fig S2A indicate all the carotenoid biosynthesis genes identified in pea aphids.

h. What is represented by blue boxes on Fig. S2?

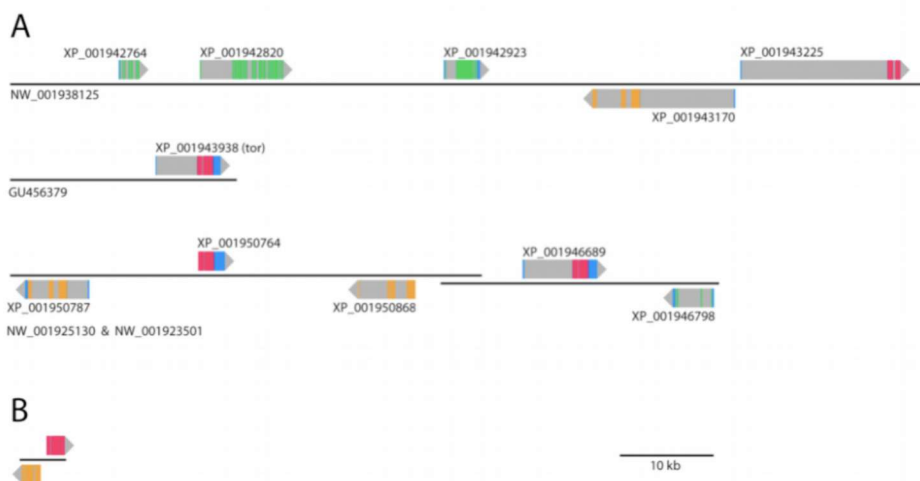
i. What is special about the location of the phytoene synthase/carotenoid cyclase genes and carotenoid desaturase genes in the pea aphid genome?

Table 1. Genes in the *A. pisum* genome with closest homology to carotenoid biosynthetic enzymes, including scaffold of origin and matching EST sequences. Similar color indicates that the gene is on the same scaffold. The 3' end of scaffold NW_001925130 overlaps with the 5' end

of NW_001923501 for 5400 base pairs, and PCR demonstrated continuity of these scaffolds. Pink row is the gene corresponding to *tor*^R and conferring red color (see text). Protein length, amino acids; ESTs are those present in GenBank, mostly from clone LSR1.

Enzyme type	Protein		mRNA		ESTs (n)	LOC_ID/ACYPID_ID	Scaffold			Gene	
	Accession	Length (aa)	Accession	Length (bp)			Accessions	Length (bp)	Start	End	Length (bp)
Carotenoid synthase/cyclase	XP_001943170	608	XM_001943135	1,981	4	LOC100161104/ACYPID002354	NW_001938125/SCAFFOLD9039	96,434	60,929	75,839	14,910
Carotenoid synthase/cyclase	XP_001950787	588	XM_001950729	2,223	9	LOC100159332/ACYPID000715	NW_001925130/SCAFFOLD17863	49,317	1,564	8,266	6,702
Carotenoid synthase/cyclase	XP_001950868	589	XM_001950833	1,770	8	LOC100164140/ACYPID005179	NW_001925130/SCAFFOLD17863	49,317	36,365	42,379	6,014
Carotenoid desaturase	XP_001943225	373	XM_001943190	1,247	3	LOC100159050/ACYPID000460	NW_001938125/SCAFFOLD9039	96,434	76,368	93,106	16738
Carotenoid desaturase	XP_001950764	528	XM_001950729	2,718	7	LOC100161380/ACYPID002604	NW_001925130/SCAFFOLD17863	49,317	19,665	22,451	2,786
Carotenoid desaturase	XP_001946689	526	XM_001946654	2,693	25	LOC100169110/ACYPID009757	NW_001923501/SCAFFOLD16397	29,128	8,574	16,388	7,814
Carotenoid desaturase	XP_001943938	510	XM_001943903	2,410	57	LOC100169245/ACYPID009883	NW_001918682/SCAFFOLD12059	31,283	1,415	8,641	7,226

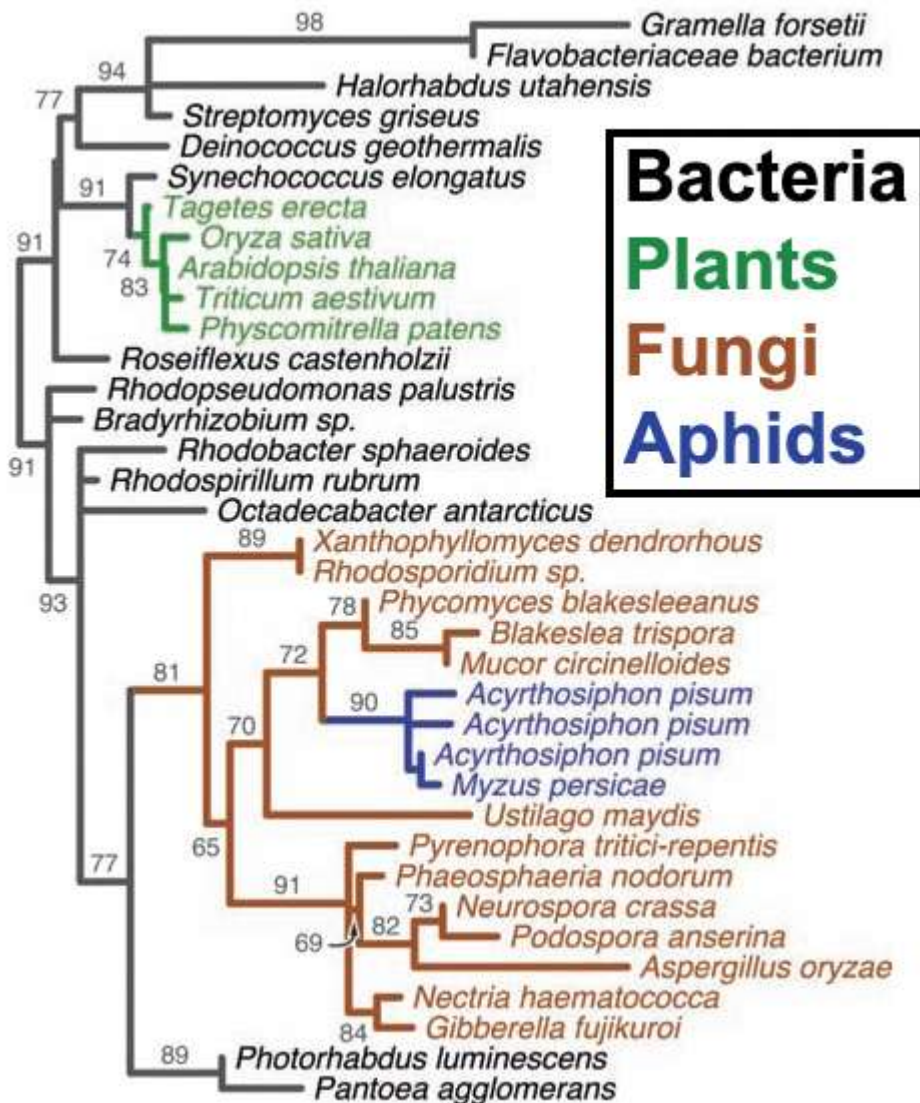
Fig. S2. Arrangement of the carotenoid biosynthetic genes (A) in the *A. pisum* genome and (B) in the genome of the fungus *Blakeslea trispora* (S12). Black lines represent genomic sequence. Genes are represented by block arrows with blue for UTR, grey for introns. Exons of carotenoid desaturase genes are red, exons of phytoene synthase/carotenoid cyclase genes are orange, and exons of other genes are green. Scaffold numbers from the *A. pisum* genome project are given, except for the region corresponding to the *tor* locus for which our sequence accession number is given.



To obtain a broader comparison of taxa, let's take one of the pea aphid sequences (one of the largest ones: ACCESSION XM_001943135) and search across animals.

g. Run a TBLASTX search with ACCESSION XM_001943135 as a query across all animals ("Metazoan") and analyse the results. This search can take a while...

All the sequences aligning to carotenoid synthase cyclase genes were retrieved and a phylogenetic tree was made:



h. Interpret the results.

i. Search NCBI Taxonomy (<https://www.ncbi.nlm.nih.gov/taxonomy/>) to check what *Myzus persicae* is.

j. Use time tree (<http://www.timetree.org/>) to estimate the time when the exchange occurred between fungi and aphids.

k. Is this an upper bound or a lower bound of the time point? What would you do to get the other bound of the time estimate?

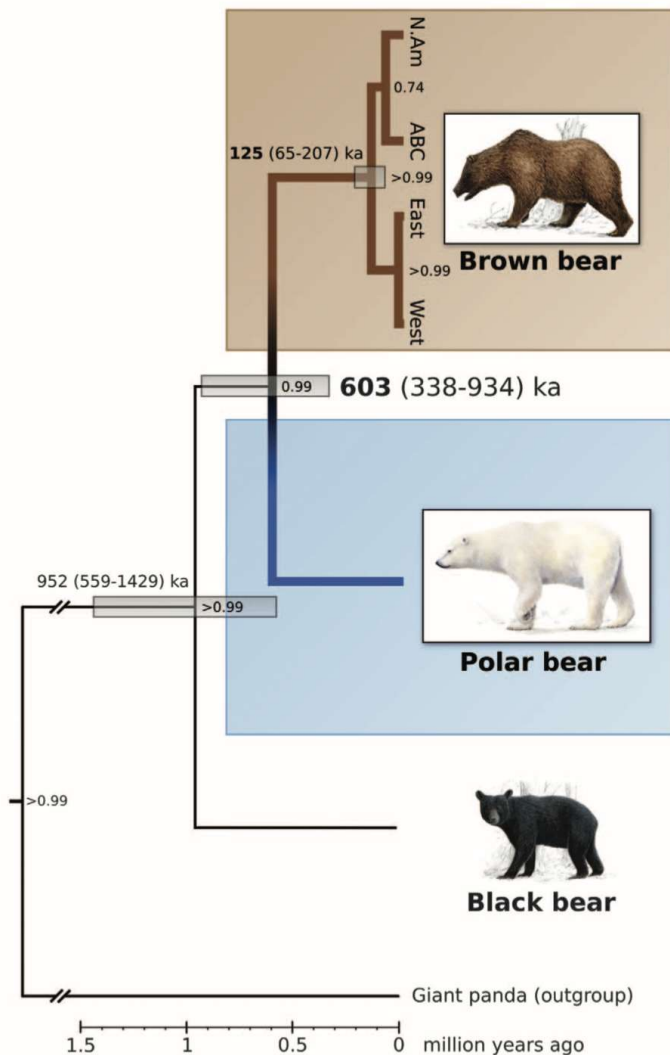
II. Polar bears

DNA was extracted from one giant panda and 45 individuals of polar, brown, and black bears. 14 independent loci (introns) were PCR-amplified and sequenced for each DNA sample. A Bayesian approach was used to reconstruct the phylogeny based on these 14x46 sequences (panel A). With high statistical support ($P > 0.99$), polar bears were recovered as a sister lineage to all brown bears.

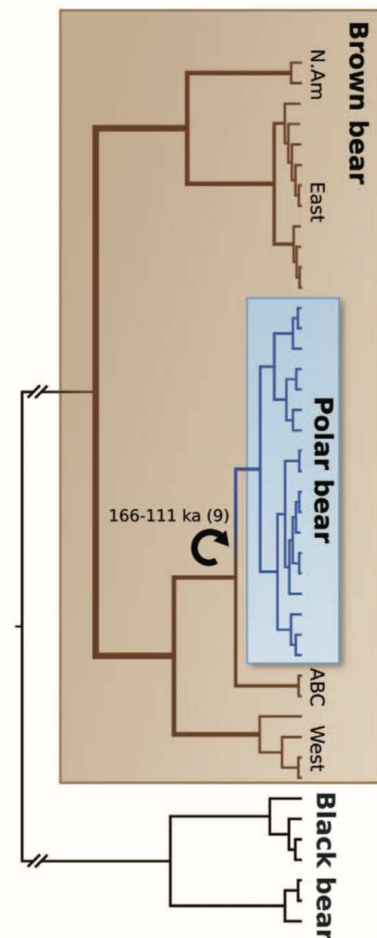
A few mitochondrial genes were also sequenced in several individuals polar, brown, and black bears and they produced another phylogenetic tree which is also highly robust statistically (panel B).

a. Interpret the two trees.

A Nuclear DNA

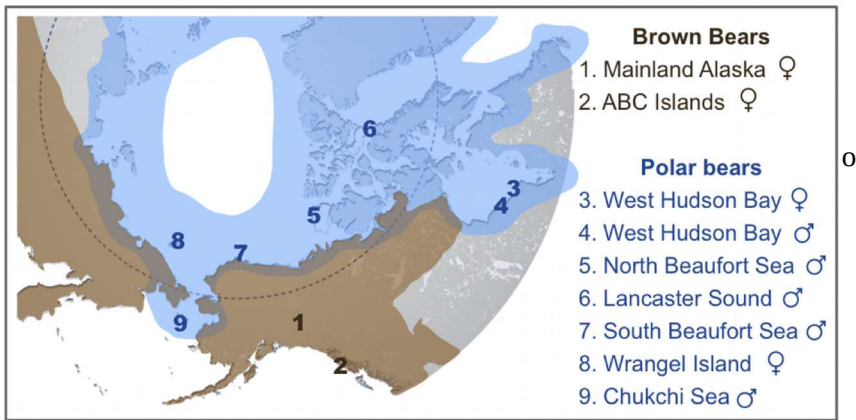


B Mitochondrial DNA



ABC = brown bears from Admiralty, Baranof, and Chichagof (ABC) Islands off the coast of Alaska .

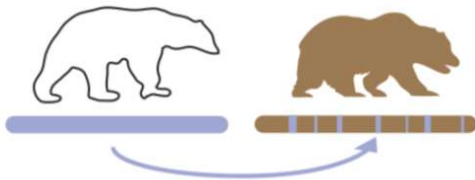
Note that the probability of joining “ABC” bears to “N. Am” bears as sister groups based on nuclear DNA (panel A) has a probability of 0.74, suggesting that the mitochondria-based scenario may be too simplistic. To try to resolve better the evolutionary history of the polar bears and the ABC brown bears, random genomic shotgun libraries were sequenced from seven polar bears, two brown bears (see map below for origin of each bear) and one black bear from Pennsylvania.



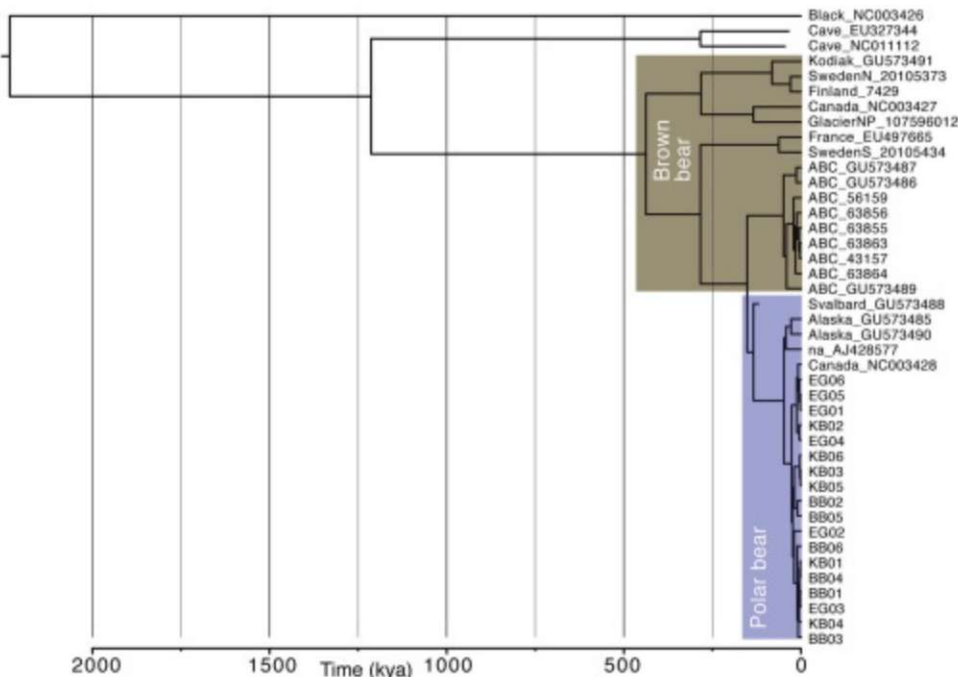
Genetic diversity was found to be lower for polar bears than for brown bears: polar bear alleles differ at about 4 sites in 10,000 whereas brown bears differ at about 17 sites in 10,000.

b. What can you infer from this observation?

To obtain more data, 79 Greenlandic polar bears and ten brown bears from Fennoscandia; mainland US; and the Admiralty, Baranof, and Chichagof (ABC) Islands off the coast of Alaska were sequenced. ABC brown bears were found to carry pieces of polar bear genomes, but not the other brown bears:



A phylogeny based on all mitochondrial data was redone and gave the same result as the first study, with polar bears being the sister group of ABC brown bears:



c. Propose a scenario for the origin of polar bears and ABC bears that matches all available data.

III. Evolution of lice

Humans can be infected by three types of lice: *Pediculus humanus capitis* (human head louse), *Pediculus humanus corporis* (human body louse) and *Pthirus pubis* (crab louse). Our ape relatives are infected by only one type of louse species.

Pediculus humanus capitis and *Pediculus humanus corporis* are very similar morphologically and can still hybridize in the laboratory although they live on different places on the human body and possess slightly different DNA sequences. DNA comparison suggest that the most recent common ancestor of *Pediculus humanus capitis* and *Pediculus humanus corporis* lived less than 70,000 years ago.

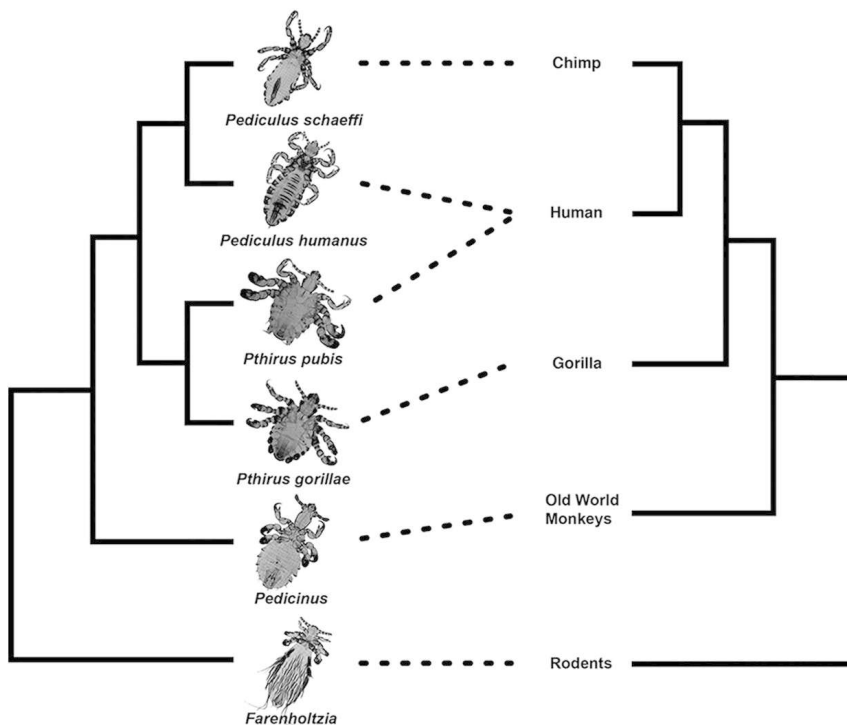
a. What can you infer about the evolution of clothes in humans?

Greater genetic diversity is found in lice from Africa compared to non-African lice.

b. Can you explain this result?

Below is a phylogeny based on both nuclear and mitochondrial genes for mammals (right) and for lice (left).

c. What can you infer? (at least two scenarios are possible)



d. Use Timetree (<http://www.timetree.org/>) to add time estimates on relevant nodes and find which scenario is most probable.

III. Cytochrome b and fungicide resistance

Quinone outside inhibiting (QoI) fungicides represent one of the most widely used groups of fungicides used to control agriculturally important fungal pathogens. QoIs inhibit mitochondrial respiration in fungi by binding to the Qo site of the cytochrome bc1 complex (which includes the protein cytochrome b), blocking electron transfer and halting ATP synthesis. QoI fungicides usually have low toxicity towards birds, mammals (including humans), and bees at limited doses. We will study the evolution of resistance to QoI.

One gene has been involved in such resistance in the grapevine downy mildew.

a. Find the case in www.Gephebase.org by searching for “grapevine downy mildew”.

b. Why are there two entries?

c. Complete the table (to compare with the lactase case that we studied in class):

Gene	<i>lactase</i>	<i>cytochrome b</i>
Type of gene	nuclear	
Trait	lactose tolerance	
Taxonomic status	intraspecific	
Presumptive Null	no	
Molecular Type	cis-regulatory	
Aberration Type	SNP	
Molecular Details of the Mutation	T14009G (rs869051967) (ss 820486563)	

Click on “cytochrome b” to see if there are other entries in Gephebase for this gene.

d. Is it correct to call QoI molecules fungicides?

To see all the mutations in cytochrome c, click on “Split Mutations”, “Submit” and then the tab “Other” (or go to <https://www.gephebase.org/search-criteria?and+Gene%20Gephebase=%5Ecytochrome%20b%5E/and+splitMutations=true>)

e. Can you find any amino acid change that evolved multiple times in cytochrome b besides Gly143Ala?

The G143A mutation confers high (complete) resistance, which is always associated with the failure of QoIs to control disease.

We will check whether the mutation G143A is also found in other species.

f. Let’s first retrieve the cytochrome b protein sequence of *Plasmopara viticola*:

Click on one of the *Plasmopara viticola* Gephebase entry (for example <https://www.gephebase.org/view-gephe/2041>), click on the Main Reference. You should then see the paper in NCBI PubMed. In the bottom right, click on “>Related Information >Protein”. Select one of the *Plasmopara viticola* protein sequence for “apocytochrome b”.

g. Check position 143 in the protein. Can you see the G143A mutation?

h. Let’s search for other cytochrome b proteins with the same G143A mutation:

Copy one of cytochrome b protein sequences of *Plasmopara viticola* harboring the G143A mutation (for example <https://www.ncbi.nlm.nih.gov/protein/ABF29522.1>). Perform a BLAST to search (choose the right

one) for similar sequences in the nr database. Click on “Alignments” and “Flat query-anchored with dots for identities” to see all the protein sequences aligned to your query. Find position 143.

i. Can you find other sequences with the G143A mutation? Which species do they correspond to? (use right-click to open the pages in other tabs)

There may be other species with the mutation but by default BLAST searches only show the first 100 target sequences. To retrieve more sequences, go back to the BLAST initial page, click on “Algorithm parameters” and select “5000” for Max target sequences. This run will take a bit longer.

j. Can you find more sequences with the G143A mutation? Which species do they correspond to?

Experimental data from *Saccharomyces cerevisiae* and *Strobilurus tenacellus* suggest that the N261D mutation in *cytochrome b* also confers resistance.

k. Use the same strategy as above to check whether the mutation N261D is also found in other species:

Let’s first retrieve the cytochrome b protein sequence in *Strobilurus tenacellus*:

Click on the *Strobilurus tenacellus* Gephabase entry (<https://www.gepibase.org/view-gepibe/2045>), click on the Main Reference. You should then see the paper in NCBI PubMed. In the bottom right, click on “>Related Information >Protein”. Select the *Strobilurus tenacellus* protein sequence.

Check position 261 in the protein. Perform a BLAST to search for similar sequences in the nr database. Click on “Alignments” and “Flat query-anchored with dots for identities” to see all the protein sequences aligned to your query. Find position 261.

Can you find other sequences with the N261D mutation? Which species do they correspond to?

l. optional: search for the G143A mutation in humans.