Cursushandleiding
Evolutiebiologie 2
2024

Table of Contents

Cursusbeschrijving

Dit vak geeft een totaaloverzicht van de evolutiebiologie. In deze cursus zal evolutie op alle biologische niveaus bestudeerd worden: op het niveau van DNA en genen, via ontwikkeling, tot morfologie, gedrag en ecologie. In de cursus proberen we die verschillende niveaus te integreren. Deze cursus is een verdieping van evolutiebiologie 1, dat je in je eerste jaar hebt gehad. Begrippen uit het eerste jaar, zoals adaptatie, seksueel dimorfisme en altruïsme worden kort herhaald.

In de eerste twee weken wordt gefocust op fylogeniereconstructie, het maken van evolutionaire stambomen. Je wordt in de (werk)colleges getraind om fylogenieën te construeren met verschillende methodes. Vervolgens ga je, met DNA sequenties die je zelf uit varens geïsoleerd hebt, een moleculaire stamboom maken.

In de derde en vierde week staat de Moderne Synthese, de klassieke basis van evolutiebiologie, centraal. Hierbij ligt de nadruk op populatiegenetica, zowel met computermodellen als met een selectie-practicum met salamanders, waarbij je zelf een voorspelling over selectie doet, SNPs genotypeert en rekent met allelfrequenties.

In de vijfde, zesde en zevende week wordt de kennis van de geschiedenis van het leven stevig uitgebreid, onder andere met een fossielenpracticum. Accenten worden tevens gelegd op veroudering, life history en de relevantie van evolutie voor de menselijke gezondheid. Evolutionaire aspecten van bijvoorbeeld het coronavirus komen aan de orde.

In week 7 wordt aandacht besteed aan evo-devo (evolutionary developmental biology): de ontwikkelingsbiologische basis van evolutionaire veranderingen. In een practicum zul je zelf de ontwikkeling van meelkevers met behulp van RNAi beïnvloeden.

Onderwerpen worden geïntroduceerd tijdens colleges en uitgediept tijdens werkcollegels/pc practica en natte practica. De werkcolleges/pc practica worden in het Sylvius in de practicum zalen gegeven. In de natte practica, ook in de practicumzalen van het Sylvius, gebruiken we moleculaire technieken, zoals DNA isolatie, PCR, sequencing en SNP genotypering, die je moet begrijpen en ook bij de tentamenstof horen. Natte practica worden in koppels van 2 studenten uitgevoerd.

We behandelen het boek *Evolutionary Analysis* van Herron en Freeman. We volgen de structuur van dit boek en dit boek helpt je dus bij het houden van overzicht. In het rooster zijn per dag de te lezen hoofdstukken, de te volgen colleges en de practica aangeven. De cursus is erg vol. Als je in tijdnood komt, focus dan vooral op de colleges, werkcolleges en practica. Daar zullen de tentamenvragen over gaan. Het boek biedt vooral houvast qua structuur en als naslagwerk.

Veel plezier en succes!

Ben Wielstra (cursuscoördinator),

Maurijn van der Zee, James France, Dick Groenenberg, Lars van den Hoek Ostende, Ken Kraaieveld, Kees Koops, Willem Meilink, Renske Onstein, Olga Paulouskaya, Martin Rücklin, Eric Snijder, Anagnostis Theodoropoulos, Manon de Visser

Phylogeny ‘werkcolleges’ & fern practicum

Before each of the three phylogeny ‘werkcolleges’ and the ‘building the fern tree’ practicum, make sure you have followed the associated lectures. Altogether there are **52** questions/assignments spread throughout the three ‘werkcolleges’ and the ‘building the fern tree’ practicum. Answers need to be compiled in a single document, to be uploaded as a PDF file in Brightspace. Students are allowed to work together in groups of 2-3. This is a particular advantage if someone cannot run a particular software or has a relatively slow computer (pick the fastest computer to run time consuming analyses). However, **every student needs to upload their own file with answers**. On your computer, create a dedicated folder where you place all your files; this is particularly relevant when during analyses many additional files will be generated by certain programs.

Installing the computer-programs for phylogeny

Dick Groenenberg, Peter Hovenkamp† & Ben Wielstra

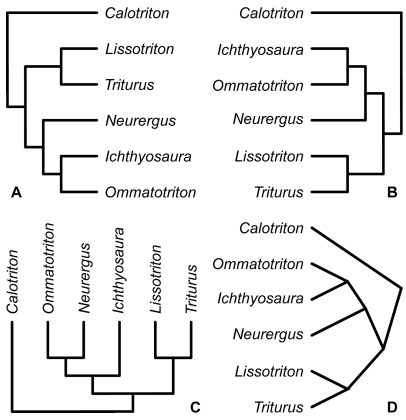
Before starting with the computer practicum, make sure that all programs are installed and working. Try to have a clean desktop at all times. Make sure your programs and data are easily accessible (i.e. know the location of installed programs, in- and output files). Create shortcuts on your desktop if you're comfortable using these. For installation of the necessary programs, follow the instructions below. They are tested for recent versions of Windows and Mac operating systems. However, do not be surprised if you encounter issues with a particular software, in that case simply collaborate with another student.

Install instructions

Werkcollege 1: Why phylogeny?

Ben Wielstra

Exercise 1 – tree thinking

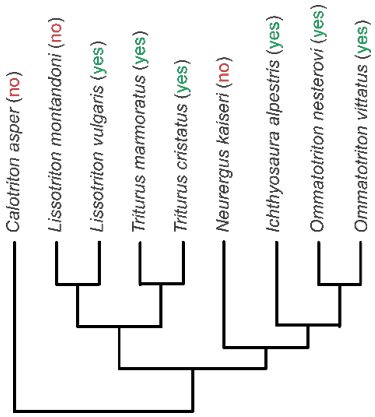


Above are four topologies showing phylogenetic relationships among six genera of European newts.

1. Which topology shows different phylogenetic relationships compared to the other three?

2. Which genus is used to root the phylogeny?

3. What is the sister taxon of Lissotriton?

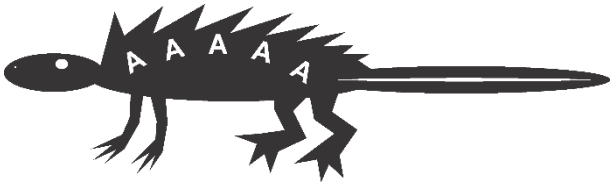


Above is a phylogeny for several newt species. For some species, the males sport a crest when in breeding attire. Other species lack a crest. 'No crest' is considered the ancestral condition, because all more distantly related newt species do not possess a crest either. Accepting this phylogeny as true, the character 'crest presence' can be *optimized*: the variation in character states in the species under study can be explained in the most economical way, requiring the least amount of character state changes to be inferred.

4. Optimize the character 'crest presence' with states 'yes' or 'no' on the phylogeny above; in other words, show where evolutionary changes happened (provide a picture/screenshot).

5. The number of steps required is greater than the theoretical minimum (which would be 1 step). How could this be?

Exercise 2 – networks and phylogenies



The fictional group of species known as 'letter newts' is composed of four species, with the illustrious names Newt A, Newt B, Newt C and Newt D.

6. How many networks are possible? Draw them all out (insert a picture).

7. On how many branches could each network be rooted?

8. How many fully bifurcating (fully resolved) topologies are possible? Draw them all out (insert a picture).

Newt * is the most closely related newt taxon and is used as an outgroup.

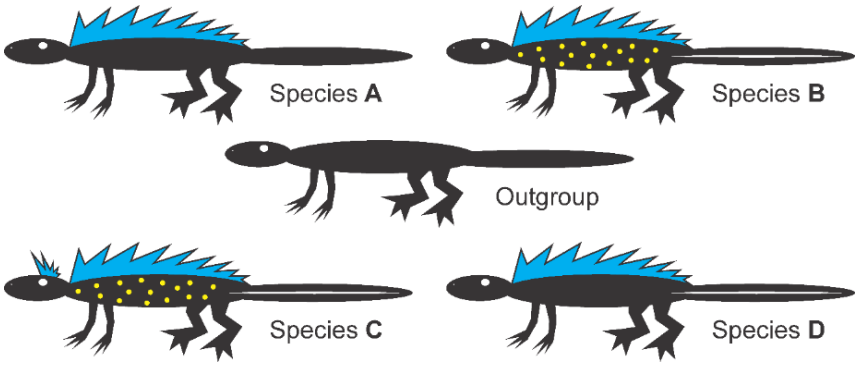
9. How many fully bifurcating topologies are possible if the outgroup is added?

The five newt taxa introduced above differ in whether they have a crest and in the colour of their bellies (which can be black or orange) as follows:

- Newt * has a crest and has a black belly
- Newt A has no crest and has an orange belly
- Newt B has a crest and has a black belly
- Newt C has no crest and has a black belly
- Newt D has no crest and has an orange belly

10. Optimize the characters 'crest presence' and 'belly colour' over all possible topologies. How many steps does each topology require and which one is most parsimonious (provide a picture)?

Exercise 3 – build your own phylogeny



Above you see life-like depictions of five hypothetical newt species. We are going to determine the phylogenetic relationships among a group of four closely related species, named species A, B, C and D, that together constitute the ingroup. A relatively distantly related species is included as outgroup and we consider the character states in the outgroup to be ancestral.

11. Summarize the available morphological data in a matrix.

12. What is the most parsimonious phylogeny? (Draw it, make a picture, and include it as a figure.)

13. Provide the most parsimonious phylogeny as a string of text (i.e. using notation with commas and parentheses).

14. Which characters are informative for solving the phylogenetic relationships within the ingroup?

15. Do species A and D constitute a clade?

16. Do species B and C form a monophyletic group?

Werkcollege 2: Morphological phylogeny

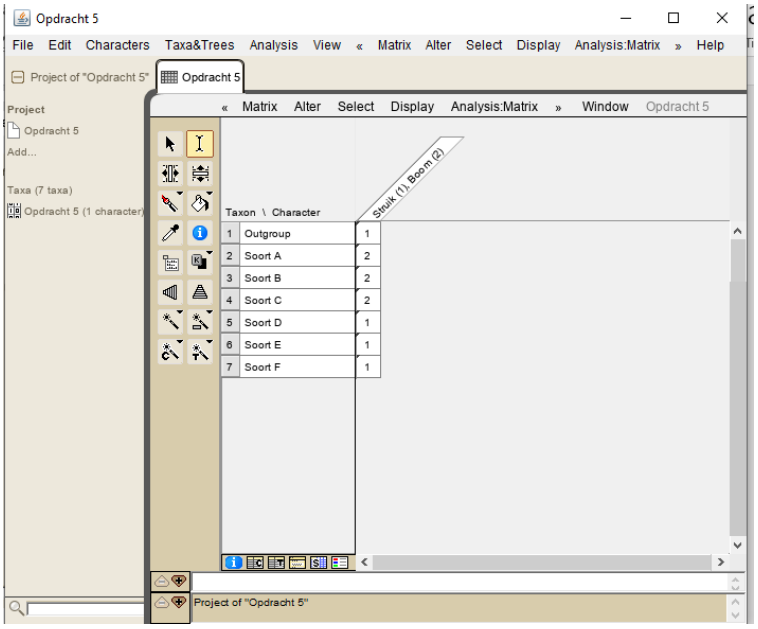
Peter van Welzen & Ben Wielstra

Exercise 4 – creating a datamatrix in Mesquite

Below you find a description of seven plants. Your job is to make a data matrix out of this description. Identify the characters and code them with discrete states. Note that not each character can be coded.

Outgroup
Struik. Blad simpel, 15-20 cm lang, rand gaaf. Kelkblaadjes 4, groen. Kroonblaadjes 4, wit. Meeldraden 8. Vrucht hard.
Soort A
Boom. Blad samengesteld, 16-19 cm lang, rand gaaf. Kelkblaadjes 4, rood. Kroonblaadjes 4, wit. Meeldraden 7-9. Vrucht zacht.
Soort B
Boom. Blad samengesteld, 16-21 cm lang, rand gaaf. Kelkblaadjes 5, rood. Kroonblaadjes 4, wit. Meeldraden 8. Vrucht zacht.
Soort C
Boom. Blad simpel, 14-18 cm long, rand gaaf. Kelkblaadjes 5, rood. Kroonblaadjes 4, wit. Meeldraden 7-8. Vrucht zacht.
Soort D
Struik. Blad simpel, 14-20 cm lang, rand gaaf. Kelkblaadjes 4, groen. Kroonblaadjes 5, geel. Meeldraden 8. Vrucht zacht.
Soort E
Struik. Blad simpel, 17-21 cm lang, rand getand. Kelkblaadjes 4, groen. Kroonblaadjes 5, geel. Meeldraden 8-9. Vrucht zacht.
Soort F
Struik. Blad simpel, 4-5 cm lang, rand getand. Kelkblaadjes 4, groen. Kroonblaadjes 5, wit. Meeldraden 8. Vrucht zacht.

Make a new datamatrix in Mesquite (you can choose to write it out on paper first). Open Mesquite (in Windows use the .jar file, on a Mac use the Mesquite_Starter_Flex.app). Go to File > New. Save your file in a dedicated folder. Tell Mesquite how many taxa are included and check 'Make Character Matrix' and press OK. In the next menu, choose 'Standard Categorial Data' and tell Mesquite how many characters you want to include. For taxa double click on individual cells and change the name (it is convenient to put the outgroup at the top of the list). For characters click on the cells and provide appropriate names and states (e.g. bladrand gaaf / getand). Now fill in the character states for each character.



Example of the making of a datamatrix in Mesquite, showing the first character coded

17. Which character did you skip and why?

18. Provide a screenshot of your final datamatrix in Mesquite.

Save your file because you will need it again later. Also, export your file > File > Export > Conservative Nexus. Provide a unique name and save in your dedicated folder. You will need this file in the next exercise.

Exercise 5 – maximum parsimony in PAUP

We will now conduct a maximum parsimony phylogeny analysis of our datamatrix in PAUP. Execute the 'Conservative Nexus' file you made with Mesquite in PAUP > File > Open > select your file. Tick the box Execute and then press the button Execute. Now we will conduct a Heuristic search: Analysis > Heuristic Search > click Max Trees and select Automatically increase by 100 and OK > go to tab Stepwise Addition and tick Random with 10 replicates > confirm that under tab Branch swapping TBR (Tree Bisection and Reconnection) is selected > OK.

19. What is shortest tree length for any of the phylogenies found? Are multiple equally long trees found and if yes how many?

We will now take a look at the best tree(s) found. Go to Trees > Describe trees > click All > Describe. For each tree you see some statistics. These also include the Consistency Index (CI) and Retention Index (RI). Both are not 1. This means there is homoplasy and there are autapomorphies.

20. Provide a screenshot of at least one of your trees and the associated statistics.

21. Are the taxa A and B consistently recovered as sister species?

First we will save our best trees, so we can use them again later in Mesquite. Go to the tab Trees > Save Trees to File > provide a unique name.

Next we will create a consensus tree based on the equally good trees we found. Go to the tab Trees > Compute Consensus > select both Strict and Majority-rule > OK.

22. Provide a screenshot of both the strict and 50% majority-rule consensus trees.

23. Are the taxa A and B recovered as sister taxa in these concensus trees?

Finally, we will determine the support for each clade using a bootstrap analysis. Go to the tab Analysis > Bootstrap/Jacknife > make sure under resampling method Bootstrap is ticked, make sure under type of search Full Heuristic is ticked > increase the number of replicates to 1000 > Continue > OK.

24. Provide a screenshot of the tree with bootstrap support values.

25. Would you accept taxa A, B and C as a monophyletic group?

Exercise 6 – analysing character evolution in Mesquite

Go back to Mesquite and open both your datamatrix that you made previously in Mesquite as well as the file with best trees that you saved in PAUP. Go to the tab File > Open File and select the file containing your datamatrix. Now you have to link your file with trees from PAUP to this datamatrix. Go to the Taxa & Trees > New Tree Window > With Trees from Source > Use Trees from Separate NEXUS File > OK > select your .tre file previously saved in PAUP and Open. You can now browse through you trees (arrows on the top left). You still need to tell Mesquite which taxon is the outgroup. Click on the icon  and then click on the branch leading to the ougroup. Now we will trace character changes. Go to the tab Analysis:Tree > Trace Character History > Parsimony Ancestral States > OK. For each character, check where they change from which state into which state in each tree. Note that when Mesquite cannot decide the character state for a particular branch, because two options are possible, that branch will have two colours. Do this for all your trees.

26. Provide a screenshot that shows a random character change in a random tree

27. Which characters express homoplasy?

28. Which character shows an autapomorphy?

29. Which character shows a synapomorphy for species A-F?

Werkcollege 3: Alignment and molecular phylogeny

Ben Wielstra

Exercise 7 – making an alignment in Geneious

Download the file "**COI_newts_fasta_raw.txt**" from Brightspace (Week 2, 13 September) and open it with a text editor. This file contains sequences of the mitochondrial COI gene for a selection of salamanders in the family Salamandridae, with the Axolotl (of the closely related family Ambystomatidae) included as an outgroup. There are errors in some of these sequences that you will need to solve below (in the program Geneious). The format of this file is the regularly used FASTA format. Notice how the sequences are represented. Each sequence has a header which start with a ">" and is followed by the species name. The sequence itself starts on a new line. After you had enough of staring at these sequences, close the file.

Open Geneious. Create a new folder titled e.g. "EB2". Drag the file "COI_newts_fasta_raw.txt" in or use the Import function in the menu. Geneious will ask you some question on how to import the sequences. If prompted, these are "Nucleotide sequences". Select "create a sequence list". On the right-hand side you can see several tabs. On the tab "General" under "Colors" make sure the second scheme is selected, which gives different bases a different colour. Under the tab "Display" select translation, so you can see the translation of the sequence into amino acids, and under "Colors" select a scheme that colours the background (e.g. the tenth, "MacClade"). Zoom in so that you can see the first 100 or so positons and can see the individual DNA letters.

30. What is the proper genetic code to use for this dataset?

Change the genetic code accordingly. Although you have changed the genetic code, there are still stop codons ("*" on a black background) in the amino acid translations.

31. Why would there still be stop codons in the sequences? More than one answer could apply.

Under "Translation" on the "Display" tab go to "Frame". Here you can select which open reading frame to display, meaning which DNA letter is treated as the first, second and third codon position when translating the sequence to amino acids. Select "Forward frames" to display all three possible open reading frames in the forward orientation.

32. At which codon positon does the open reading frame start and why?

One species is not in the same orientation as the others. Rather, it is displayed as 'reverse complement'. This can happen if your reverse primer is interpreted as forward primer and the other way around. In effect, the sequence is in the opposite direction and the complementary base pairs are interchanged (A for T and C for G). E.g. the reverse complement of ACCTGAG is CTCAGGT.

33. Which species is displayed as reverse complement and how did you recognize this?

Correct the orientation of the species in 'reverse complement' by selecting it and reverse complementing it (so basically a second time). There are several ways to accomplish this, for example with a right mouse click. Make sure you select "Reverse complement selected sequences".

For one species initially it was not clear what nucleotide a particular position was. It is encoded by an IUPAC code that reflects this uncertainty (you can find a list of the IUPAC codes e.g. here: https://www.bioinformatics.org/sms/iupac.html).

34. Which species has an IUPAC code reflecting uncertainty in its sequence, which position is affected and what potential nucleotides could be here?

35. Does this ambiguity influence the amino acid translation? What codon position is affected and does that surprise you?

A closer look has revealed the uncertain position to be an A. Change it accordingly (if prompted, allow editing).

For one species an extra letter has erroneously been inserted in the sequence, resulting in a frame shift.

36. Which species has an extra letter inserted in its sequence, which letter is inserted and which position is affected?

Remove the extra letter in the sequence.

37. After these corrections, are there non-synonymous substitutions left in the alignment?

All the errors in the data matrix have now been corrected and we are ready to export the file so it can be used to build a phylogeny. File > Export > Selected documents (or Ctr+Shift+E) and if it asks to save intermediary steps press “Save”. Save as FASTA sequences/alignment but change “raw” to “corrected” in the file name. Press “Proceed” and “Ok” when prompted.

Exercise 8 – molecular phylogeny with MrBayes

The next thing we need to do is convert your corrected FASTA file to a NEXUS format including commands for the program MrBayes. For this we will use the program Mesquite, which you used before, as intermediary. When opening your file (File > Open File), Mesquite prompts you to provide the format of the file you are opening. Choose FASTA (DNA/RNA). Next it will immediately ask you to save it as a NEXUS file. Do so. Now export the file (File > Export) as a “NEXUS file for MrBayes”. Mesquite will now show you a ‘MrBayes block’ it will attach to your NEXUS file. This block contains a set of commands that MrBayes will use (see below). Export the file, without changing the MrBayes block, with a unique name (e.g. include “MrBayes”). Now you can close Mesquite.

Open your NEXUS file with a text editor. Notice how the sequences are represented this time. Each sequence has a name and is followed by a white space. The sequence itself is in the same line. The file consists of a data block, which includes details on the dimensions of your file (the number of taxa and characters), the data type (DNA in this case) and the actual data matrix.

38. How long is your alignment?

The data block is followed by the MrBayes block, currently including the default list of commands added by Mesquite. If no outgroup is defined in the MrBayes block, the program automatically assumes the first taxon is the outgroup. It is possible to define multiple partitions for a sequence (multiple genes and multiple codon positions for example), but we treat our sequence as a single partition. The lset, unlink and prset line define the model of sequence evolution used (in this case the General Time Reversible model with a proportion of invariable sites and a gamma-shaped distribution of rates across sites). Programs exist that determine which model best fits your data, but we skip this step here.

39. Will MrBayes pick the right taxon as outgroup?

The mcmcp line provides the commands for the execution of the analysis. The Markov chain Monte Carlo analysis will run with four chains (three heated and one cold) which is actually the default. Under "ngen" the number of generations for which the analysis will run is defined and "samplefreq" determines how often the chain is sampled ("printfreq" is the frequency with which the state of the chains is displayed on the screen). "savebrlens" means branch lengths are saved alongside the sampled tree topology. Because the tree simulation starts at a random place in 'tree space' the first trees sampled are likely to be far from the optimum and these are discarded as 'burn-in', "burninfrac" sets what proportion of trees is removed. The program TRACER allows you to determine if you ran enough generations and how many trees to discard as burn-in, but we skip this step here.

40. How many trees will be removed as burn-in under this unedited MrBayes block?

In the MrBayes block the number of generations is set to 10 million. This requires too much computing time for our purposes, so change this to 1 million manually. Also, to be able to read the consensus phylogenetic tree produced by MrBayes in Mesquite, add " conformat=simple" (note the space) between "sumt" and ";". Do not change any other settings and save your file.

Make sure that the NEXUS file you created is in a dedicated folder and not, say, on your desktop, because MrBayes is going to create a lot of files. Open MrBayes (the .exe file is in the "bin" folder). Type "exe" followed by a space, drag your file into the MrBayes window, and press enter (if the path is too long MrBayes will give an error, if that is the case reduce the path length; also, do not use spaces in the path). Now your analysis starts. MrBayes will give you an estimate of how much time remainss until it is finished. How much time the analysis takes depends on the specs of your computer. After MrBayes has finished it will display some output statistics and some (poorly formatted) trees. We will not look at the output in MrBayes directly. Close the program by typing "quit" followed by an enter or just press the cross.

Next, we will open the Bayesian phylogeny we just created in Mesquite. Open Mesquite and open the consensus tree that MrBayes created, which is a file ending in "con.tre". Under "interpretation" select "posterior probability". Next, Taxa&Trees > New Tree Window > With Trees from Source. Choose "Stored Trees".

41. Make a screen shot of your tree and paste it in your answer sheet

42. Which genus is sister to the genus Salamandra?

43. Is the genus Ommatotriton monophyletic or paraphyletic

44. What is the support value of the clade encompassing the genus Triturus

45. Does this analysis provide a good insight on the relationships within the Salamandridae family and why?

Building the fern phylogeny 1: DNA extraction and PCR practicum

Barbara Gravendeel, Dick Groenenberg, Peter Hovenkamp†, Klaas Vrieling & Ben Wielstra

Outline

The goals of the 'Building the fern phylogeny' practicum are to gain experience with wet-lab techniques (DNA extraction, PCR, gel-electrophoresis), sequence editing and assembly, BLAST searching and phylogenetic analyses.

For this practicum material has been collected by Roderick Bouman and Gerda van Uffelen from ferns growing in the Hortus Botanicus Leiden. Each students will receive a numbered sample. The material has been identified, but species names will be released later. With this sample, each of you will follow these steps:

1. Perform a DNA extraction.
2. Check the DNA extract on a gel.
3. Perform a PCR on the DNA extract.
4. Submit the PCR product for sequencing in forward and reverse direction.
5. Assemble forward and reverse sequences in Geneious.
6. Perform a BLAST search with the obtained contig sequence for preliminary identification.
7. Download the full set of sequences (with outgroup), align them and perform phylogenetic analyses.
8. Use Mesquite to map a character (sporangium) and draw conclusions about its evolution.

Background

Using sequence data for identification ("DNA barcoding")

For DNA barcoding, one or more short parts of the DNA of organisms under study are sequenced as an identification method to quickly identify it as belonging to a particular species. The method is based on a relatively simple concept: most eukaryote cells contain mitochondria and mitochondrial DNA with a relatively fast mutation rate, which results in significant variance of sequences between species and a comparatively small variance within species. A region of around 650 bp of the mitochondrial gene, known as cytochrome c oxidase I (COI), was therefore initially proposed as a potential barcode for animal species identification. DNA barcodes are currently collected all over the world in thousands of molecular laboratories and deposited in an internationally accessible web-based database: the NCBI GenBank (Baker et al. 2003; Harris, 2003). For plants, the most widely used markers are on the chloroplast DNA, which behaves in many ways as the mitochondrial DNA.

Pitfalls of DNA barcoding

DNA barcoding has been met with mixed feelings, ranging from enthusiastic endorsement to vociferous opposition. For example, many stress the fact that DNA barcoding does not provide reliable information above the species level. Others resent what they see as a gross oversimplification of the science of taxonomy. And some suggest that recently diverged species might not be distinguishable on the basis of their DNA sequences. Exploratory studies have shown that about 96% of eukaryotic species surveyed can be detected with barcoding - though most of these would also be resolvable with traditional means - the remaining 4% pose problems which can lead to error rates that are unacceptably high when relying on DNA barcoding alone. Empirical evidence is also accumulating that some DNA barcoding markers can be multicopy genes. This means that multiple nonfunctional gene copies exist that are sometimes co-amplified when using them as DNA barcoding marker. Care must be taken to use homologous copies when applying these markers since the use of non-homologous copies can result in a wrong identification (Song et al. 2008).

Challenges to solve

Ideally, DNA barcoding is used alongside traditional taxonomy so that problematic cases can be identified and errors can be detected. Non-cryptic species can generally be resolved by either traditional or molecular taxonomy without ambiguity. However, more difficult cases might only be resolved by a combination of approaches. Another factor to consider is that, because (almost) all mtDNA and cpDNA genes are maternally inherited, hybridization will lead to misleading results. Solutions also need to be found when type specimens anchoring species names cannot be used to generate a DNA barcode. In these cases, type specimens consist of a drawing, water colour or photograph, without any link to a biotic specimen. Even with biotic specimens available to extract and amplify DNA from, though, most of the global biodiversity is still unknown, so DNA barcoding can only hint at the existence of new taxa, but not delimitate or describe them (DeSalle, 2006).

From DNA sequence data to phylogenetic trees

Obviously, the goal of DNA barcoding is identification of species, not the reconstruction of genealogical relationships. DNA barcoding typically relies on obtaining the same DNA fragment (one marker) for all species. For the inference of genealogical relationships it is best to rely on as many markers as possible (nowadays even genome data: phylogenomics), because markers can differ in evolutionary rates and/or histories. Phylogeny reconstructions based on DNA sequence data from one marker will reflect the gene tree for that marker, not necessarily the species tree. An example of the latter phenomenon is incomplete lineage sorting. Consider the following situation where the lineages of two genes (blue and red) are shown with the known phylogeny of species ABC

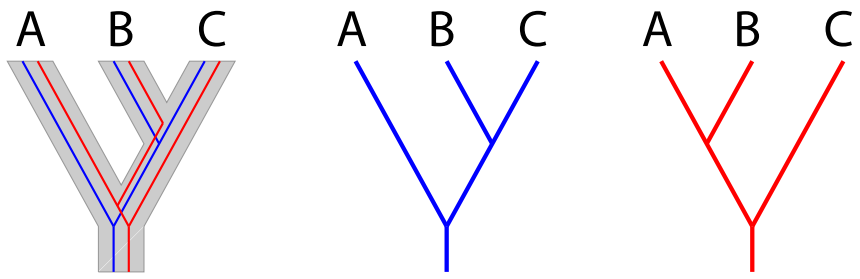


Figure 1: Species tree (left) versus gene trees (blue and red).

When reconstructing phylogenies based on sequence data, we hope or expect (based on shared common ancestry) that most gene trees will be congruent with the species tree, as is the case for the blue gene. In case of the red gene the coalescense between C and AB pre-dates the split between A and BC, i.e. for the red gene the bifurcation between C and AB in is older than that of the organism lineage BC. A gene tree based on sequence data for only the red gene would result in the false conclusion that A and B are sister species. In other words, there is a discordance between the species tree and the red gene tree.

In this practicum we will construct gene trees based on a single marker, as approximation of the phylogenetic tree.

Introducing ferns and sporangium evolution

The fern samples are selected to give a broad overview of ferns. On basis of the selected marker (rbcL) we have good hope that we can reconstruct the main relationships between the sampled species (see, e.g., Hasebe et al. 1995). On basis of this, we will try to interpret the evolution of sporangia in ferns. Sporangium diversity in ferns includes a number of distinct forms (fig. 2). The form generally considered primitive is the Eusporangium, which contains numerous spores and which has a thick wall that is mainly passive when the sporangium opens (Hovenkamp et al. 2009). The Leptosporangium contains a limited number of spores (mostly 64), and has a thin wall with an annulus composed of thick-walled cells, a structure that actively contributes to the dispersal of the spores. A rough classification of different types of leptosporangia again distinguishes asymmetric sporangia, in which the annulus is oblique, outside a plane of symmetry, and symmetric sporangia, which have a strictly vertical annulus in the plane of symmetry. This is the fern sporangium that is most widely known. The list of fern names that will be distributed also lists the annulus types.

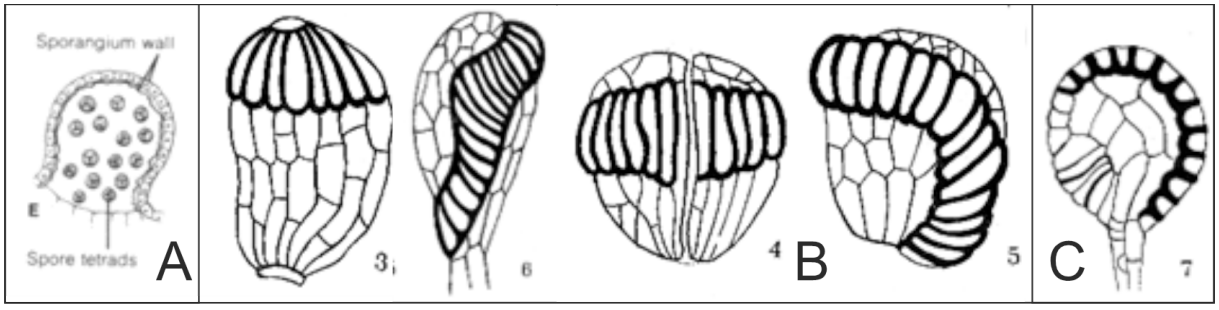


Figure 2 A: Eusporangium, B: various forms of asymmetric leptosporangia, with oblique annulus C: symmetric sporangium with vertical annulus

On basis of the reconstructed fern phylogeny, what can you infer about the evolution of the fern sporangium? Is the Eusporangium indeed primitive? What is the most derived form? What might have been the impact of the changes on fern evolution?

References

Baker, C. S., Dalebout, M. L., Lavery, S. & Ross, H. A. 2003. www.DNA-surveillance: applied molecular taxonomy for species conservation and discovery. Trends in Ecology & Evolution 18: 271-72.

DeSalle, R. 2006. Species discovery versus species identification in DNA barcoding efforts: response to Rubinoff. Conservation Biology 20(5): 1545-1547.

Harris, D. J. 2003. Can you bank on GenBank? Trends in Ecology & Evolution 18: 317-19.

Hasebe, M., P. G. Wolf, K. M. Pryer, K. Ueda, M. Ito, R. Sano, G. J. Gastony, J. Yokoyama, J. R. Manhart, N. Murakami, E. H. Crane, C. H. Haufler and W. D. Hauk. 1995. Fern Phylogeny Based on rbcL Nucleotide sequences. Amer.Fern J. 85: 134-181.

Hovenkamp, P. H., R. Van Der Ham, G. A. Van Uffelen, M. Van Hecke, J. A. Dijksman and W. Star. 2009. Spore movement driven by the spore wall in an eusporangiate fern. Grana 48(2): 122-127.

R.M.R. & Strobeck, C. 2008. Beyond mtDNA: nuclear gene flow suggests taxonomic oversplitting in the little brown bat (*Myotis lucifugus*). Canadian Journal of Zoology 86(7): 700-713.

PPG1. 2016. A community-derived classification for extant lycophytes and ferns. Journal of Systematics and Evolution 54(6): 563-603.

Song, H., Buhay, J.E., Whiting, M.F. & Crandall, K.A. 2008. Many species in one: DNA barcoding overestimates the number of species when nuclear mitochondrial pseudogenes are co amplified. Proceedings of the National Academy of Sciences of the USA 105 (36): 13486-13491.

DNA extraction ferns

Before you start

Read each step of the protocol carefully and keep in mind where the DNA is. You are going to use the DNA for further experiments so do not throw it away after it has been put on a gel! DNA extraction is a very important step in most molecular biology work. The success of subsequent experiments considerably depends on the quality and quantity of the DNA samples! The names of the plants are not yet revealed when you are given the sample (1 per student couple). So carefully label your samples.

Label the samples

A problem encountered every year is lousy (and/or lazy) labeling of tubes. It seems such a simple thing but it turns out to be the hardest part of every procedure in molecular techniques. Therefore, be sure to label your tubes properly. What should be written on a tube?

1) Sample name (e.g. F12)

2) group number (each couple has a unique group number, e.g. 11).

This results in a label in the format F12_11 (fern sample F12 processed by student group 11). **Never give your sample a new name, use the original name!** What is in the tube is not only your business, but it should also be clear to everybody else.

DNA Extraction from Plant Tissues

We use a commercial kit for the DNA extraction - the DNeasy Plant Mini Kit from QIAGEN. The DNeasy DNA extraction kit allows rapid and efficient extraction of high quality DNA from many different plant species and tissue types. The following protocol is modified from the handbook in the kit. Figure 3 shows a flow chart of the procedure.

Important points before starting

All centrifugation steps are carried out in a microcentrifuge at room temperature (15–25°C). The extraction will take circa 2 hours, so make sure you have time to finish it when you start

1. Check that all required buffers and enzymes are available (AP1, RNase A, P3, AW1, AW2 and AE)
2. Make sure that a heating block has been pre-heated to 65°C.

Personal safety measures

Be sure to wear safety goggles when working with liquid nitrogen. Be very careful as liquid nitrogen can cause severe burn wounds.

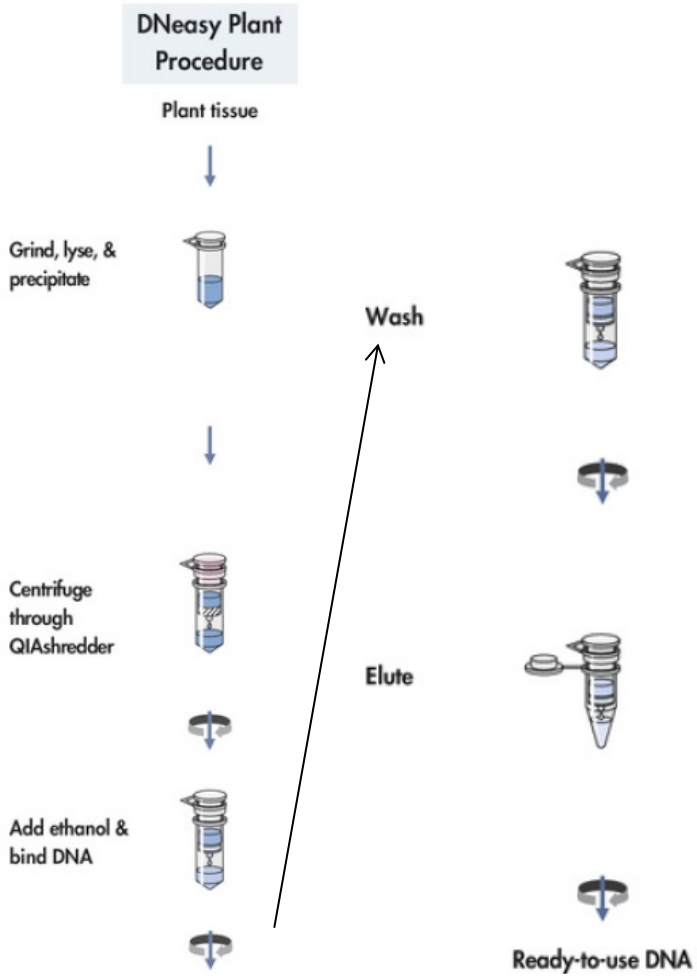


Figure 3. Schematic representation of the DNA extraction procedure.

Anti-contamination measures

Be careful not to cross-contaminate samples! For instance, don't reuse pestles without sterilizing them first. Keep in mind that DNA template molecules compete with each other (endogenous vs. exogenous) in subsequent PCRs (a small amount of DNA extract from a fresh specimen can easily outnumber the amount of template molecules from an aged (museum) specimen). Wear gloves to protect your sample from human DNases.

Important! Once samples are frozen, the samples should not be defrosted before the grinded material is transferred into the extraction buffer! When the samples are defrosted degradation of the DNA will start.

Extraction steps

3. Heat extraction buffer AP1 of the DNeasy kit to 65°C.
4. Pre-cool a 1.5 ml Eppendorf tube (eppie).
5. Cut up about 1 cm^2 leaf and add it to the eppie. Fill the eppie halfway with liquid nitrogen and cool the pestle in liquid nitrogen before grinding.
6. Start by gently crushing the leaf material immediately. Start grinding when the liquid nitrogen is almost evaporated. When you arrive at a fine powder (circa 10~20 s grinding after the liquid nitrogen has evaporated) add 400 µl of the warm Buffer AP1 of the DNeasy kit. Add 4 µl of RNase A (100 mg/ml).

   (If defrosted before the sample is in the (safe) extraction-buffer, phenolic compounds will irreversibly bind to the DNA (your sample may turn brown), which inhibits DNA extraction and later on PCR. In the extraction buffer the pH is high to prevent breakdown of your DNA sample, it contains EDTA (or a similar chelating agent) to immobilize Mg2+ ions that function as a cofactor for DNases. The RNase is added to hydrolyse/breakdown RNA.)
7. Vortex the sample well and incubate it for 10 min in a heating block at 65˚C, mix two or three times during the incubation by inverting the tube (lyses of cells). Make sure that after vortexing all plant material is mostly in the buffer. If part of it stays on the lid, hit the tube onto the table with moderate force.
8. Add 130 µl of Buffer P3 of the DNeasy kit and mix well. Incubate the sample for 5 min on ice (sample may be left on ice for longer than 5 min without damage). Centrifuge samples for 10 min at 12000 rpm.

   (At low temperatures detergents, proteins, and polysaccharides are precipitated. Through centrifuging these precipitates and cell debris are collected at the bottom of the tube. The DNA stays in the solution.)
9. Transfer the supernatant (with 1000 µl pipet) to the QIAshredder Mini spin column (lilac column) placed in a 2 ml collection tube. Centrifuge for 5 min at 12000 rpm. The DNA stays in the solution and therefore ends up in the collection tube.

   (This column is used to remove precipitates and cell debris that are still floating in the solution)

10. Pipette the flow-through fraction (about 450 µl) from the collection tube to a new 1.5 ml tube without disturbing the cell-debris pellet.

    (The QIAshredder Mini spin column removes most precipitates and cell debris, but a small amount may pass through and sometimes form a pellet in the collection tube. Be careful not to disturb this pellet.)

11. Directly pipette 1.5 volumes (about 675 µl) of Buffer AW1 onto the clear lysate and mix well by gently inverting tube.

12. Pipette 650 µl of the mixture (including any precipitate that may have formed) into the DNeasy Mini spin column (white column) placed in a 2 ml collection tube. Centrifuge for 1 min at 8000 rpm. Discard the flow-through and reuse the collection tube for step 12!

    (The DNA is binding to the silica membrane in the column. Note that you did not use all the solution from step 10. The collection tube can contain at maximum 650 µl and therefore not all the liquid from step 10 is brought on the column. The remainder of the liquid of step 10 is brought on to the DNeasy Mini spin column in the next step)

13. On the (same) DNeasy Mini spin column from step 11 add the remaining fluid from the sample tube from step 10 (so **DO NOT** start with a new DNeasy Mini spin column here!!!) and centrifuge for 1 min at 8000 rpm. Discard the flow-through and the collection tube!

    (DNA binds to the membrane of the DNeasy Mini spin column.)

14. Place the DNeasy Mini spin column into a new 2 ml collection tube. Add 500 µl Buffer AW2. Centrifuge for 1 min at 8000 rpm. Discard flow-through and reuse the collection tubes!

    (In this step the DNA is still bound to the membrane. The AW2 buffer is used to wash contamination away.)

15. The washing step is repeated. Add again 500 µl Buffer AW2 to the DNeasy Mini spin column. Centrifuge for 5 min at 12000 rpm to dry the membrane. Remove the DNeasy Mini spin columns from the collection tube carefully so the column does not contact with the flow-through. Put the DNeasy Mini spin column to a clean 1.5 ml tube. Discard the flow-through and the collection tube. To be sure all the ethanol is evaporated leave the sample for 10 minutes at room temperature.

    (It is important to dry the membrane of the DNeasy Mini spin column since residual ethanol may interfere with subsequent reactions.)

16. Pipette 100 µl Buffer AE (elution buffer) directly onto the membrane of the DNeasy Mini spin column. Leave the sample for 5 minutes at room temperature (15 ~ 25 °C). Centrifuge for 1 min at 8000 rpm. Keep spin column in 1.5 ml tube.

    (The pH of the AE buffer allows the DNA to dissociate from the silica membrane. As a consequence the DNA is dissolved again in AE solution. To remove additional DNA from the membrane this step is repeated with a smaller volume of AE in the next step.)

17. Pipette again 50 µl Buffer AE (elution buffer) directly onto the membrane of the DNeasy Mini spin column. Leave the sample for 5 minutes at room temperature (15 ~ 25 °C). Centrifuge for 1 min at 8000 rpm.
18. Make sure the DNA samples are properly labelled with (1) sample name and (2) group number in the format "F12_11"and store them properly (the practicum organizers will provide a tray). The total volume in the 1,5 ml tube should be about 150 µl.

    (If the DNA is used on short notice (as in your case) and/or frequently, store it in refrigerator (about 4 ˚C) to avoid repeated freezing and thawing (because this degrades the DNA). Long-term storage of the DNA sample is best done at −80 ˚C, or, if not possible, at −20 ˚C. To ensure uncontaminated and high quality DNA remains, typically most of the DNA extraction is directly stored frozen long-term, while an aliquot of the DNA extraction is kept separately in the fridge for frequent use.)

DNA on gel

Checking the quantity and quality of the extracted DNA

The extracted DNA will be checked for quality and quantity with electrophoresis on an agarose gel. Good DNA will leave the slots completely and will show a clear distinctive band on the gel. Degraded DNA will give a smear on one side of the DNA-band (away from the slots). The quantity of DNA will be estimated with electrophoresis by comparing your extracted DNA with that of the molecular weight ladder. It is necessary to estimate the quantity of the DNA to prevent large differences between the samples. Differences between the samples will cause difficulties for PCR and might eventually interfere with the interpretation of the patterns of DNA fingerprints: e.g. is the band vague because of a lack of DNA or due to too much DNA being present?

The DNA will be colored with SYBR safe. SYBR safe is caught in the DNA helix (intercalary) and is fluorescent under UV-radiation. The intensity of the fluorescence is proportional to the quantity of DNA. After comparing your sample with that of the DNA ladder the quantity of DNA in your sample can be estimated. Then DNA samples can be diluted to specific concentrations for different experiments.

For the Fern samples 2-4 gels are made by the practicum organizers for the for all samples in each practicum room.

(As a complement, spectrophotometers like the NanoDrop (http://www.nanodrop.com) can be used to quantify the concentration of DNA samples.)

Molding the gel

1. Weigh 1 g agarose in an Erlenmeyer flask.
2. Add 100 ml 0.5 x TBE buffer.
3. Add 10 µl SYBR safe (which binds to DNA and can be visualized under UV light – the fact that it binds to DNA underscores that gloves must be worn for protection!)
4. Heat for approximately 1 min in the microwave. Shake the solution afterwards very carefully as it might start boiling unexpectedly upon shaking (superheating).
5. Make sure that the solution is completely clear (all agarose dissolved)
6. Let the solution cool down to approx. 55 ~ 60 °C (5 min), check with a thermometer.

   (At 35 ~ 40 °C, agarose coagulates. If the solution is warmer than 60 °C it will damage the plastic electrophoresis tray)
7. Pour the cooled solution in the template (the electrophoresis tray).
8. Immediately place the (2) comb(s) to make the slots on the marked positions. Make sure that all teeth are put into the agarose.
9. Let the gel coagulate for 15 ~ 20 min.

   (Do not move the template to prevent tearing)
10. After coagulation of the gel, fill the tray with (used) electrophoresis-buffer (0.5 x TBE) to just above the gel.

    (This prevents the gel from dehydrating)
11. Carefully take the comb out of the gel.

Preparation of the DNA samples

1. Apply 2 µl loading buffer on a small piece of parafilm.

   (The loading buffer contains glycerol and a dye. By diluting your sample in glycerol it will become heavier than the electrophoresis-buffer and when you apply the DNA to the slots it will sink and stay in there. Because the loading buffer contains a dye it allows you to see the progress of the electrophoresis)
2. Use a P10 or P20 to pipet 4 µl DNA extract and mix this on a small piece of parafilm with the 2 µl of loading buffer (take a clean pipet-point for each sample). Save the DNA extract as it will be used for PCR later on.
3. Write your (1) Sample name and (2) group number in the format “F12_11” on the gel form in the right place!
4. Load the 6 µl of DNA/loading buffer-solution in the slots. (Clean, after application, the point of the pipette by pipetting the electrophoresis-buffer a few times, then apply with the same pipette point the next sample. Why is contamination not a problem here?)

5. After all samples are loaded, apply 6 µl of Generuler ladder per row on the gel. This range of concentrations of DNA can later on be used to estimate the size and quantity of DNA.
6. Connect the power source and electrophorese for circa 30 ~ 45 min at 100~120 V. Because DNA is negatively charged, it moves from the negative electrode (black) to the positive electrode (red). Make sure the gel-tray is positioned in the correct direction, otherwise DNA will run out of, instead into, the gel. Run the gel covered by aluminium foil because the SYBR safe stain degrades when exposed to light.

   (By following the front of the loading buffer you can see the progress of the electrophoresis)

When the front of the loading buffer has tracked approximately 3 cm, the electrophoresis can be terminated. The DNA can now be visualized under UV light (by a practicum organizer). The gel picture will be made available on Brightspace when ready, you will need it to answer the question below.

46. Provide a picture of your DNA sample on gel. Did you obtain good results?

PCR amplification chloroplast barcoding gene rbcL

Primers

The primers used with this practicum are:

Forward primer PLANT_(M13F-rbcLaf)_F:

5' TGTAAAACGACGGCCAGTATGTCACCACAAACAGAGACTAAAGC 3'

Reverse primer PLANT_(M13R-rbcLar)_R:

5' CAGGAAACAGCTATGACGTAAAATCAAGTCCACCRCG 3'

(Primers from CBOL Plant Working Group (2009) A DNA barcode for land plants. Proc Natl Acad Sci USA 106: 12794–12797).

Underlined are the M13 sequences. The M13 sequence does not match with the plant DNA but it will be the part where the sequencing primer will anneal.

PCR mastermix

Per table prepare a PCR master mix. So if you have 6 samples prepare a mix for 7 samples. The 1 sample is extra is to make sure that there is enough mix to fill out the 6 reactions and the remaining mix is used for a blank (i.e. a negative control).

Below the recipe is given for 1 one sample. Make a calculation and show it to one of the practicum organizers before you start pipetting.

Each PCR reaction mix contains the following ingredients:

Reaction mix for 1 sample:	
Sterile water (MilliQ)	12.9 µL
CoralLoadPCR Buffer (10×))(contains 1.5mM Mg Cl2)	2.5 µL
dNTP's (2.5 mM each)	2.5 µL
Forward primer (10umol/µl)	2.0 µL
Reverse primer (10umol/µl)	2.0 µL
Mg Cl2 (25mM)	1.0 µL
Taq (5 units/µl)	0.1 µL
Total	23.0 µL

Thaw all ingredients before use (except the Taq polymerase that is already fluid). **At the moment the ingredients are defrosted put them immediately on ice.**

Put also your mix on ice during preparation. Make the PCR mix in a 1.5 ml Eppendorf tube. Keep the order of the pipetting as in the scheme above (i.e. from large to small volumes). Start with pipetting the water and end with the Taq polymerase. Mix the sample well by pipetting up and down with a 1 ml pipette (not vortexing!) but take care that no foam is formed. If foam is formed during pipetting centrifuge shortly.

One couple uses the remainder of the PCR mix to prepare a blank. For the blank 2 ul of water is added instead of the DNA.

Amplification by PCR

1. Label the PCR tube correctly, write on lid *and* side, in the format "F12_11". Proper labeling is essential for good results.
2. Add 23 µl of reaction mix to a 0.2 ml PCR tube.
3. Add 2 µl of your DNA extraction and pipet up and down to mix homogeneously.
4. Gently tab the PCR tubes so that all fluids are at the bottom. Confirm that your sample name, and group number are on the tube (in the format "F12_11"). Place your PCR tube in the 96 well rack on ice provided by the practicum organizers and note where in this rack your sample is on the associated sheet (in the format "F12_11").
5. The practicum organizer will the PCR tubes in the PCR machine. The PCR program is as follows:

   a) Denature (15 min at 95° C)
   b) Denature (30s at 95° C)
      Annealing (30s at 51° C)
      Extension (1min at 72° C)
   c) Repeat cycle b 35 times
   d) Final Extension (5 min at 72° C)
   e) pause step at 12° C

Checking PCR success

The quantity and quality of the PCR product will be checked on an (precast) eGel by using 4 µl of the PCR product by the practicum organizers. The gel picture will be made available on Brightspace when ready, you will need it to answer the question below.

47. Provide a picture of your PCR product on gel. Did you obtain good results?

Submitting the product for sequencing

Depending on the results, PCR-products will be selected for further purification and cycle sequencing. Purification and cycle sequencing will be outsourced.

Building the fern phylogeny 2: processing and analyzing your sequences

Ben Wielstra

We will start with the c. 30 minute online lecture 'Sequence analysis: Make your own trees!' by Dick Groenenberg, which provides further background of today's practicum and instructions on the different steps. You can re-watch (parts of) this lecture during the day when required.

Prepare your sequence

With Geneious, create a consensus sequence for the data returned from BaseClear. All sequences can be downloaded as a zipped file **"Zipped_fern_sequences"** from Brightspace. Pick out your own sequences as well as other sequences with the same sample name. In this practicum, material from each fern was (often, not in all cases) extracted multiple times, yielding multiple forward and reverse sequences with the same sample name. Proceed with all sequences for your sample. Import the sequences into Geneious and discard 'bad' sequences (messy electropherogram with many 'double' peaks or very small unreadable peaks). Proceed with the 'good' sequences only and make a contig.

48. Provide a screenshot of the first c. 100 bp of your 'contig'

Identify your sequence

Do a Blast-search to confirm the identification of your consensus sequence. Confirm that the closest matches are with taxa in the same group (ferns) as your sample. If not, your sequence is most likely not from the sample, but from a contamination (e.g. a fungus or bacterium). If the sequence appears to be correct, see which species is the best match, and use that name as (provisional) identification of your sample. Save the sequence as a FASTA file and make sure that the name reflects their identification (also include the sample code and group number; use underscores instead of spaces). Now compare your identification with the list available on Brightspace (file called **"Varenlijst"**).

49. Provide your own sequence in Fasta format with your identification resulting from the BLAST-search. Motivate your choice for this identification. Is it close to the 'official' identification/real name?

Align all sequences

Download the file **"Unaligned_fern_consensus_sequences"** from Brightspace. This file contains unaligned consensus sequences for the samples by all students. These consensus sequences have already been compiled by the course administrators, so students can immediately continue with the next exercise and do not have to wait for each other's sequences. Open the file in Geneious and align the consensus sequences (either manually or with one of the build in algorithms). Note that not all sequences are of equal length, i.e. there is missing data. While it is possible to trim the alignment so that all sequences are of equal length, this would also mean removing parts of longer sequences that potentially contain phylogenetic information, so we skip this step. Export the file as Fasta format as you learned before ('Werkcollege 3') under a recognizable name. Simplify the names to just the taxon name in a text editor (alternatively this could be done in Geneious before exporting the file).

50. What is the length of your alignment?

Bayesian phylogenetic inference

Convert your alignment into a MrBayes file using the skills you learned in 'Werkcollege 3'. Remember to make sure MrBayes uses the right outgroup (this is the species *Huperzia phlegmarioides*, which has sporangium type A). Execute the file in MrBayes and wait until the analysis is done (e.g. have lunch, read a book).

Character tracing

When MrBayes is finished, open the file with extension **con.tre** in Mesquite. Open the tree in Mesquite: Taxa&Trees > New Tree Window > With Trees from Source > Stored Trees > OK. Now add the character sporangium-types as an additional character: Characters > New Empty Matrix > Standard Categorical, 1 character called 'sporangium-type' > OK. Fill in the cells using the character states for sporangium type provided with the fern names (see file **"Varenlijst"** on Brightspace). Finally, trace the sporangium type over the tree (as you learned in 'Werkcollege 2').

51. Provide a screenshot of your phylogeny with the sporangium character traced

52. Which changes have taken place during fern evolution?

Computerpracticum: Selection

Part II: Mechanisms of Evolutionary Change

Case Studies in Evolution

SELECTION AND MUTATION AS MECHANISMS OF EVOLUTION

by Jon C. Herron, University of Washington

Introduction

The purpose of this case study is to help you develop an intuition about how selection and mutation cause evolution. Answers to the questions need to be compiled in a single PDF document and **uploaded in Brightspace by each student individually**. You will use a software simulation of an evolving population to analyze the examples discussed in Chapter 6, and to answer a variety of questions concerning changes in the frequencies of alleles. Once you are familiar with the simulation program, you can use it to answer questions of your own. For example, in Chapter 7 we will look at evidence suggesting that the CCR5-Δ32 allele is only about 700 years old in European populations. You can use the simulation program to estimate the strength of selection that must have been required to cause the Δ32 allele rise from a frequency of virtually zero to a frequency of 0.1 to 0.2 in less than 30 generations.

To complete the case study you will need the program AlleleA1. You can download AlleleA1 (for Windows or MacOS) from the Evolutionary Analysis website or run it in your browser. AlleleA1 simulates evolution at a single locus in an ideal population. The locus has 2 alleles: A1 and A2. AlleleA1 allows you to enter parameters controlling selection, mutation, migration, drift, and inbreeding. The program then plots a graph showing the frequency of allele A1 over time. Each generation's frequency is calculated from the previous generation's frequency, according to the equations described in Chapters 5 and 6. Small boxes in the lower portion of the AlleleA1 window allow you to enter and change the parameters for the simulation. The tool palette has buttons that allow you to run the simulation, clear the graph, reset all parameters to their default values, print your graph, and quit. More details on using AlleleA1 can be found in the software manual on Brightspace and under the Help menu in AlleleA1.

Exercises

Hardy-Weinberg Equilibrium

1. After fiddling with the simulation program to see how it works, restore all parameters to their default settings. The default settings encompass initial frequencies of 0.5 for both alleles, and the assumptions of no selection, no mutation, no migration, no genetic drift, and random mating.

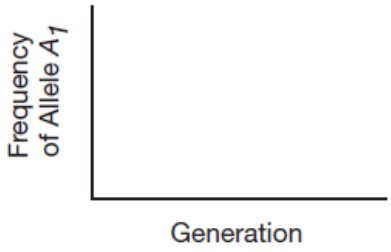
   Run the simulation to verify that under these conditions the allele frequencies do not change. Try different values for the starting frequency of allele A1. Does your experimentation verify that any starting frequencies are in equilibrium so long as there is no selection, no mutation, no migration, and no drift?

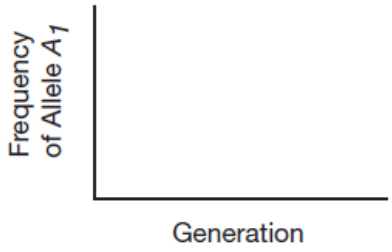
Selection as a mechanism of evolution

2. There are three boxes that let you set the fitnesses for the three genotypes. The fitnesses allow you to play with the effects of selection (that is, differences between the genotypes in survival or reproduction). Setting the values to 1, 0.8, and 0.2, for example, is equivalent to specifying that for every 100 individuals of genotype A1A1 that survive to reproduce, 80 individuals of genotype A1A2 survive, and 20 individuals of genotype A2A2 survive.

   a) Predict what will happen if you set the fitnesses of A1A1, A1A2, and A2A2 to 1, 0.8, and 0.2, respectively. Then run the simulation. Was your prediction correct? Explain

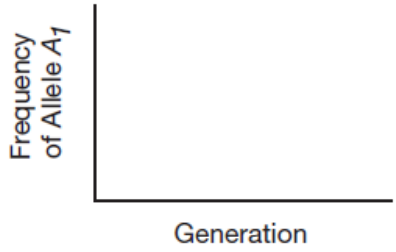
Prediction: What actually happened: Explanation:

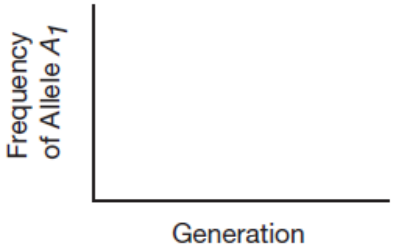




   b) Now set the initial frequency of allele A1 to 0.01, and the fitnesses to 1, 1, and 0.99. What happens when you run the simulation? Why? Now try fitnesses of 1, 1, and 0.95. Can you explain the difference?

Fitnesses of 1, 1, and 0.99: Fitnesses of 1, 1, and 0.95: Explanation:





   c) Look at Figure 6.14 in the textbook (page 209). In the research depicted in the figure, researchers raised experimental populations of fruit flies on food spiked with ethanol, and monitored the frequency of the AdhF allele over 50 generations. This also lets you know the frequency of AdhS. AdhS encodes a version of the alcohol dehydrogenase enzyme that breaks down ethanol at only half the rate of the version encoded by AdhF. The starting frequency of AdhS was about 0.65 in both experimental populations; the ending frequency was about 0.1 in one population and about 0 in the other.

      We will use AlleleA1 to estimate the strength of selection against the AdhS allele during this experiment. Let A1 represent the AdhS allele. Set the starting frequency of A1 to 0.65. Set the number of generations to 50. (To change the number of generations, use the popup menu at the lower right corner of the graph. Press on the small button with the black triangle, then select the number of generations you want.) Try different combinations of fitnesses for the three genotypes. Find a combination that reproduces the pattern of change over time in Figure 6.14 (You may want to show AdhF instead of AdhS; it is possible to change what is displayed on the y-axis using the black triangle). What combination of fitnesses works best? What do these fitnesses represent in terms of the relative survival (or reproductive success) of the three genotypes?

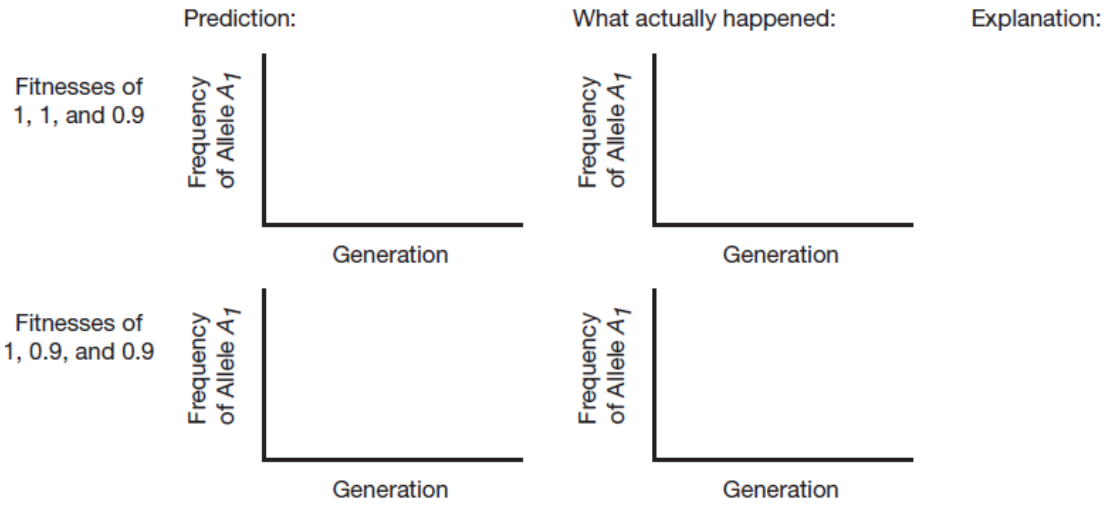
d) Reread Changes in the Frequency of the CCR5-Δ32 Allele Revisited, on page 214, and look at the graphs in Figure 6.17.

In each scenario depicted in the figure, we made our prediction based on the assumption that the fitness of +/Δ32 heterozygotes is equal to the fitness of +/+ homozygotes. In reality, +/Δ32 heterozygotes may have somewhat higher fitness than +/+ homozygotes.

Use AlleleA1 to explore which, if any, of the three predictions is strongly affected by allowing +/Δ32 heterozygotes to have higher fitness than +/+ homozygotes. Describe your results. Selection on recessive and dominant alleles

3. Restore all parameters to their default values, then set the initial frequency of allele A1 to 0.01.

a) Predict what will happen when you try fitnesses of 1, 1, and 0.9 and then check your prediction. Now predict what will happen when you try fitnesses of 1, 0.9 and 0.9, and check your prediction. Were your predictions correct? Try to explain what happened. (Hint: Reread Selection on Recessive and Dominant Alleles on pages 129-131. Try to reproduce Dawson's predictions in Figure 6.19. Then consider this question again).



b) In Question 3a, when was allele A1 dominant (with respect to fitness) and when was it recessive? Which will increase in frequency more rapidly when favored by selection: a rare recessive allele, or a rare dominant allele? Why?
(Hint: Try running various combinations of initial frequencies and fitness values in AlleleA1).

c) Which rises to a frequency of 1.0 more rapidly under selection: a common recessive allele, or a common dominant allele? Why?

Selection via eugenic sterilization

4. Similar to early 20th century eugenicists, imagine a single locus at which there is a gene controlling strength of mind:
A2 is the allele for normalmindedness, A1 is the allele for feeblemindedness and A2 is dominant over A1.
Imagine, as Henry H. Goddard (1914) did, that allele A1 has a rather high frequency of say 0.1.

a) Using pencil and paper, what is the frequency of feebleminded individuals in the population according to the Hardy-Weinberg equilibrium principle?
If we had a population of 1000 individuals, how many would be feebleminded? How many would be carriers for feeblemindedness? How many would be homozygous normal-minded?

b) If a eugenic sterilization law were universally enforced, such that all feebleminded individuals were sterilized before reaching sexual maturity, what would be the fitnesses of the three genotypes? Explain.

c) Using pencil and paper, what would be the frequency of allele A1 after a single generation of eugenic sterilization. (Use the numbers you calculated in part a, and assume that every nonsterilized individual makes exactly 10 gametes).
What is the total number of gametes? What fraction carry allele A1? What would be the frequency of feebleminded individuals?
How effective is eugenic sterilization at reducing the frequency of feeblemindedness?

d) Use AlleleA1 to predict the long-term effect of eugenic sterilization on the frequency of the allele for feeblemindedness.
For example, could feeblemindedness be eliminated within 20 generations? Why or why not? How long is 20 human generations in years? What do you think a eugenicist would conclude from this simulation? What else could be done to eliminate feeblemindedness?

Selection on homozygotes and heterozygotes

5. In the 1950's, biologists Terumi Mukai and Allan Burdick (1959) discovered that their laboratory population of fruit flies harbored a genetic locus with interesting effects on viability (that is, survival). The locus has two alleles, which we will call V (for viable) and L (for lethal). Individuals with genotype VV survive, whereas individuals with genotype LL die before reaching adulthood.
Mukai and Burdick established two separate populations of flies in which the initial frequency of allele V was 0.5. They propagated both populations for 15 generations, and monitored the frequency of the V allele.

a) Assuming that genotype VL has the same fitness as genotype VV, use AlleleA1 to predict what will happen in Mukai and Burdick's experiment. Is your prediction consistent with Mukai and Burdick's own expectation that the frequency of the viable allele would quickly rise toward 1.0?

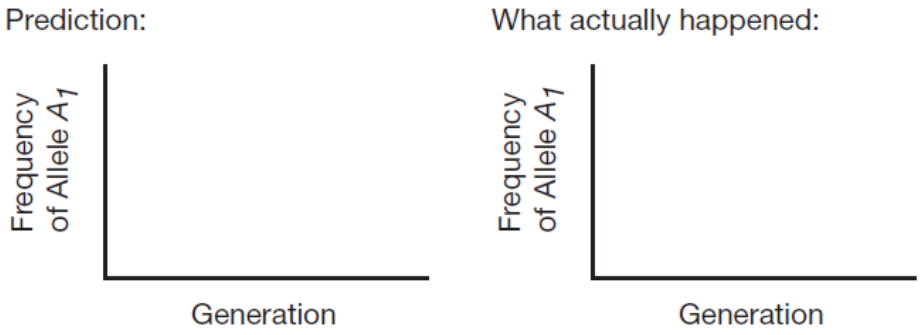
b) The actual result is shown by the red symbols in Figure 6.21 on page 221: The frequency of allele V rose, but only to a frequency of about 0.79.
Mukai and Burdick next established two populations in which the initial frequency of the viable allele was 0.975. The result for these populations is shown by the blue symbols in Figure 6.21: The frequency of allele V dropped to about 0.79. Using AlleleA1, set the initial frequency of A1 to 0.5. Experiment with different fitnesses for the three genotypes, always making sure that the values you choose are consistent with what you already know about alleles V and L (what only information do you have on VV and VL versus LL until now?). Can you find values that cause the frequency of allele A1 to rise to an equilibrium at 0.79?

c) Now set the initial frequency of A1 to 0.975. When you run the simulation, does the frequency of A1 fall to an equilibrium at 0.79? Continue to play with the simulation until you find a combination of fitnesses that works.

d) Based on your experiments, state a hypothesis that explains the behavior of Mukai and Burdick's fly populations.

Mutation as a mechanism of evolution

6. There are 2 boxes in AlleleA1's window that let you play with the mutation rate. One controls the rate at which copies of A1 turn into A2's; a mutation rate of 0.001 means that each generation one out of every thousand A1's turns into an A2. The other box controls the mutation rate in the other direction. Note that the mutation rate should be a number between 0 and 1 (why?). If you enter a number outside this range you will get weird behavior. Return all parameters to their default values, then set the mutation rates to 0.0001 and 0. Predict what will happen



Were you correct? For any real gene a mutation rate of 0.0001 would be extraordinarily high. How effective is mutation, by itself, as a force of evolution?

Mutation-Selection Balance

7. Consider the case of spinal muscular atrophy. Spinal muscular atrophy is a neurodegenerative disease characterized by weakness and wasting of the muscles that control voluntary movement. It is caused by recessive loss-of-function mutations in a gene on chromosome 5 called telSMN (SMN stands for "survival motor neuron").
   a) Using AlleleA1, return all parameters to their default values. Let A2 represent the normal allele of telSMN, and let A1 represent a loss-of-function allele.
   Brunhilde Wirth and colleagues (1997) estimate that the fitness of affected individuals is about 0.1. Set the fitnesses to 0.1, 1, and 1. What is the frequency of the knockout allele after 500 generations? Why?

   b) The actual frequency of knockout alleles for telSMN in populations of European ancestry is about 0.01. One hypothesis for the maintenance of this frequency is that new knockout alleles are continuously created by mutation.
   With fitnesses of 0.1, 1, and 1, how high does the mutation rate from A2 to A1 need to be to achieve an equilibrium frequency of 0.01 for allele A1?

8. Wirth and colleagues measured the actual mutation rate in the telSMN gene. It is high-- about 0.00011. Do you think a balance between mutation and selection is an adequate explanation for the persistence of telSMN knockout alleles at a frequency of 0.01? Explain.

9. Now consider the case of cystic fibrosis. Cystic fibrosis is a recessive genetic disease caused by loss-of-function mutations in the CFTR gene. Affected individuals suffer chronic respiratory infections that ultimately cause severe lung damage. Let A2 be the normal allele (C) and A1 the mutant allele (c).

    a) Until recently, very few cc individuals survived long enough to reproduce. Return all parameters to their default values, then set the fitnesses to 0, 1, and 1. What is the frequency of the c allele after 500 generations? Why?

    b) The actual frequency of the c allele is about 0.02 in European populations. One hypothesis for the maintenance of this frequency is that new copies of the c allele are continuously created by mutation. With fitnesses of 0, 1, and 1, how high does the mutation rate from A2 to A1 need to be to achieve an equilibrium frequency of 0.02 for allele A1?

    c) The actual rate of mutations creating new c alleles is about 0.00000067. Is a balance between mutation and selection a plausible explanation the maintenance of the c allele at a frequency of 0.02? If not, develop an alternative explanation and use AlleleA1 to demonstrate that it is plausible. See page 237 of the book for one researcher's alternative hypothesis and test.

Literature Cited

Goddard, H. H. 1914. Feeble-mindedness: Its causes and consequences. The Macmillan Company, New York.

Mukai, T., and A. B. Burdick. 1959. Single gene heterosis associated with a second chromosome recessive lethal in Drosophila melanogaster. Genetics 44: 211-232.

Wallace, B. 1963. The elimination of an autosomal lethal from an experimental population of Drosophila melanogaster. American Naturalist 97: 65-66.

Wirth, B., T. Schmidt, et al. 1997. De novo rearrangements found in 2% of index patients with spinal muscular atrophy: Mutational mechanisms, parental origin, mutation rate, and implications for genetic counseling. American Journal of Human Genetics 61: 1102-1111.

Computerpracticum: Drift and Migration

Part II: Mechanisms of Evolutionary Change

Case Studies in Evolution

MIGRATION AND GENETIC DRIFT AS MECHANISMS OF EVOLUTION

by Jon C. Herron, University of Washington

Introduction

This case study will help you develop an intuition about how migration and genetic drift cause evolution. Answers to the questions need to be compiled in a single PDF document and **uploaded in Brightspace by each student individually**. You will use a software simulation of an evolving population to analyze examples discussed in Chapter 6, and to answer a variety questions concerning changes in the frequencies of alleles. Once you are familiar with the simulation program, you can use it to answer questions of your own. For example, How often does a mildly deleterious allele drift to fixation in populations of different sizes? To complete the case study you will (again) use AlleleA1.

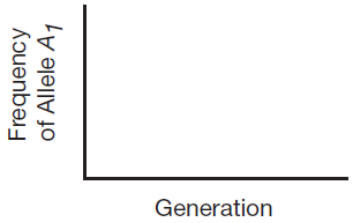
Migration as a mechanism of evolution

1. AlleleA1 uses the one-island model of migration described on pages 248-253 of the book. The simulation tracks the frequency of allele A1 in an island population.
   The parameter called Fraction of migrants each generation determines the number of individuals that move from the mainland to the island every generation, as a fraction of the island population.

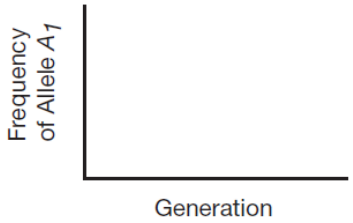
   For example, setting the parameter to 0.1 means that each generation ten percent of the individuals in the island population are new arrivals from the mainland. The parameter called Frequency of A1 in the source pop'n determines the frequency of allele A1 on the mainland (and thus among each generation's migrants).

   a) Click on the Reset button to restore all parameters to their default values.
   Predict what will happen when you set the fraction of migrants each generation to 0.01 and the frequency of A1 in the source population to 0.8. Then set the parameters to these values and run the simulation. If your prediction was not correct, try to explain the difference between what you expected and what actually happened.

Prediction:



What actually happened:



Explanation:

   b) Leave the frequency of A1 in the source population at 0.8, and try setting the fraction of migrants each generation to 0.05, then 0.1.

c) Try several different values for the both the fraction of migrants and the frequency of allele A1 in the source population.

d) Based on your experiences in parts a, b, and c, summarize what migration from the mainland does to the frequency of A1 on the island. How long does it take for migration to exert its influence? How effective is migration as a mechanism of evolution?

Migration and selection

2. Imagine that allele A1 is deleterious for individuals living on the island, such that the fitnesses of genotypes A1A1, A1A2, and A2A2 are 0.9, 0.95, and 1.

a) If there is no migration, what is the frequency of A1 after 500 generations? Why?

b) In accord with what you saw in 2a, set the starting frequency of A1 to 0. Remember that this is the frequency in the island population. Now imagine that although it is deleterious on the island, allele A1 is beneficial on the mainland, such that A1 is fixed in the source population (that is, its frequency is equal to 1).

c) What is the island frequency of A1 after 500 generations if the fraction of migrants each generation is 0.0001? 0.001? 0.01? 0.1?

Fraction of migrants	**Ending frequency of A1**
0.0001	
0.001	
0.01	
0.1	

d) In the scenario you investigated in 2b, how high does the migration rate have to be for migration to overwhelm selection in controlling the frequency of A1 on the island?
If selection against A1 on the island were stronger than we assumed in this example, would migration be less likely to overwhelm it?

3. Reread Empirical research on migration as a mechanism of evolution on pages 251-253 in the text. Richard King and colleagues studied an example of selection and migration in Lake Erie water snakes that is similar to the scenario you investigated in question 2.
On islands in Lake Erie, the allele for banded coloration is deleterious. The allele is fixed on the mainland, however, and migrants move from the mainland to the islands each generation.
Box 7.2 on pages 252-253 describes an algebraic analysis by King and Lawson (1995) of the equilibrium in the island population between selection and migration.

We can use AlleleA1 to do a similar analysis. Reset all parameters to their default values. Let allele A1 be the dominant allele for the banded pattern, and A2 be the recessive allele for the unbanded pattern. Set the starting frequency of A1 to zero, and the frequency of A1 in the source population to one.

a) To reflect King and Lawson's best estimates, set the fitnesses of A1A1, A1A2, and A2A2 to 0.84, 0.84, and 1. Set the fraction of migrants each generation to 0.01.
What is the frequency of allele A2 (the allele favored on the islands) after 500 generations?

b) To reflect King and Lawson's high-end estimate (strong selection, little migration) set the fitnesses of A1A1, A1A2, and A2A2 to 0.78, 0.78, and 1, and set the fraction of migrants each generation to 0.003.
Now what is the frequency of allele A2 after 500 generations?

c) Finally, to reflect King and Lawson's low-end estimate (weak selection, much migration) set the fitnesses of A1A1, A1A2, and A2A2 to 0.90, 0.90, and 1, and set the fraction of migrants each generation to 0.024.
What is the frequency of allele A2 after 500 generations?

d) The actual frequency of the unbanded allele in the island population is 0.73.
How well did our model perform at predicting this result? Can you think of ways to modify our model to make it more realistic? If so, try them out with AlleleA1.

Genetic drift as a mechanism of evolution

4. We have so far used AlleleA1 to simulate evolution in populations of infinite size. In reality, of course, populations are finite.
Is evolution in finite populations different from evolution in infinite populations?
Return all parameters to their default settings. Set the number of generations to 15 (use the popup menu to the right of the graph's horizontal axis). Now play with populations of finite size. For example, set the population size to 10 or 20 individuals and run the simulation several times.
What happens? Why? Does the same thing happen every time? Why or why not? (For help answering these questions, reread A Model of Genetic Drift on pages 254-256.)

5. How much does genetic drift change with population size? Return all parameters to their default settings. Set the number of generations to 100. Set the graph line mode to multiple, and the graph line color to auto.
Now investigate the power of genetic drift at different population sizes:
A. Set the population size to 4 and run the simulation several times.
B. Clear the graph. Set the population size to 40 and run the simulation several times.
C. Clear the graph. Set the population size to 400 and run the simulation several times.
   a) Compare your results for parts A, B, and C to graphs a, b, and c in Figure 7.15 on page 261. Why are the results different for populations of different sizes?

6. Conservation biologists generally consider genetic diversity to be a good thing. That is, populations are more likely to escape extinction if there are several alleles present for each gene.

   a) What does drift do to the genetic diversity in a population as the population nears extinction?

   b) Reset all parameters to their default values. Note that the starting frequencies for alleles A1 and A2 are 0.5. Roughly how big does the population have to be for the chances to be reasonably good that both alleles will persist for 500 generations?

The random fixation of alleles

7. Consider the fate of a rare allele in a small population.

   a) Reset all parameters to their default values, then set the starting frequency of A1 to 0.005 and the population size to 100. Run the simulation several times.
   What usually happens? Why?

   b) Now make the rare allele beneficial (and dominant) by setting the fitnesses of genotypes A1A1, A1A2, and A2A2 to 1.1, 1.1, and 1.0. Run the simulation several times.
   What usually happens? Why?

   c) How strong does selection in favor of the rare dominant allele have to be before the allele has a reasonable chance of becoming fixed in the population instead of lost?

   d) How strong would selection have to be if the rare allele were recessive instead of dominant?

Drift, mutation, and selection

8. How big must a population be before a mildly advantageous allele will become fixed as rapidly as it would in a population of infinite size?
   Restore all parameters to their default values. Set the starting frequency of A1 to zero, and both mutation rates to 0.00001. Make A1 beneficial by setting the fitnesses of genotypes A1A1, A1A2, and A2A2 to 1.04, 1.02, and 1. Set the number of generations to 1000.

   a) Set the graph line color to black, and run the simulation once with the population size set to infinite. How long does it take for allele A1 to be created by mutation and carried by natural selection to fixation?

   b) Now set the graph line mode to multiple, the graph line color to red, and the population size to 10. Run the simulation several times. What typically happens?

   c) Set the graph line color to orange and the population size to 100. Run the simulation several times. What typically happens?

   d) Set the graph line color to green and the population size to 1000. Run the simulation several times. What typically happens?

   e) Finally, set the graph line color to blue and the population size to 10000. Run the simulation two or three times. What typically happens?

   f) How large does a population have to be before a mildly advantageous allele will become fixed as rapidly as it would in a population of infinite size?
   How strongly does the answer depend on the strength of selection and the mutation rate?

Drift, selection, migration, and genetic diversity

9. In question 6, we noted that conservation biologists consider genetic diversity to be a good thing, and found that genetic drift can reduce genetic diversity and potentially hasten the extinction of small populations. We now revisit these issues by investigating a particular scenario in more detail.

Imagine that allele A1 is maintained in a population by heterozygote advantage, with the fitnesses of genotypes A1A1, A1A2, and A2A2 equal to 0.2, 1.0, and 0.8.

a) What is the equilibrium frequency of A1 in an ideal population? (You can answer this question either by using the formula derived in Box 6.7 on pages 222-223, or by using AlleleA1.)

b) Once the frequency of allele A1 has reached its equilibrium value, what are the frequencies of the three genotypes? (You can answer this question either by using the Hardy-Weinberg equilibrium principle, or by using AlleleA1.)

c) Once the frequency of allele A1 has reached its equilibrium value, what is the mean fitness of the population? (To calculate the mean fitness, multiply the fitness of each genotype by its frequency, then sum the results; see Box 6.3 on page 208).

d) Using AlleleA1, first reset all parameters to their default values. Then set the starting frequency of A1 to 0.2, the fitnesses of A1A1, A1A2, and A2A2 to 0.2, 1.0, and 0.8, and the population size to 20. Run the simulation several times. What usually happens?

e) After A1 has been lost due to genetic drift, what is the mean fitness of the population? By how much has drift reduced the mean fitness? If this effect were multiplied across several loci, could drift substantially increase the chance that our small population will go extinct? Why or why not?

f) Use AlleleA1 to investigate the effect of introducing a single migrant individual into our small population each generation, where the migrant comes from a large population in which the frequency of A1 is 0.2. (What value for the Fraction of migrants each generation parameter reflects a single individual joining a population of 20?)
Could the introduction of migrants maintain enough genetic diversity in the population to ameliorate the effects of genetic drift? Compare the scenario you have investigated in this question to the case of the Florida panther discussed in section 7.5 on page 297 of the book.

Literature Cited

Buri, P. 1956. Gene frequency in small populations of mutant Drosophila. Evolution 10:367–402.

King, R.B., and R. Lawson. 1995. Color-pattern variation in Lake Erie water snakes: The role of gene flow. Evolution 49:885–896.

Computerpracticum: Quantitative Genetics

!!THIS MORNING'S LECTURE ON QUANTITATIVE GENETICS IS ESSENTIAL BEFORE YOU START!!

We will work with the computer program Bugsville. You can run this in your browser via this link or you can download a stand alone version from Brightspace. The software manual on Brightspace also covers Bugsville. Answers to the questions need to be compiled in a single PDF document and **uploaded in Brightspace by each student individually**.

1. Assessing the variation in spot number

Sketch—don't copy in detail—the histogram showing the distribution of spot number in your starting Left Field population. Record the average number of spots.

2. Estimating the heritability of spot count

Record your data on parents and offspring for the matings you set up in a table (example below). When you feel you have an accurate estimate of the heritability of spot count, paste a screenshot of your scatterplot with regression line. Record your estimate for the heritability of spot count.

Family Number	# Spots Parent 1	#Spots Parent 2	Midparent (average spots)	# Spots Offspring 1	# Spots Offspring 2	Midoffspring (average spots)
1						
2						
3						
Etc.						

3. Set up a selection experiment and calculate the selection differential

Provide a screenshot of a histogram showing the distribution of spot counts among your selected parents. Record the average number of spots among your selected parents. Calculate and record the selection differential for your selection experiment.

4. Predict the response to selection

Multiply your estimate of the heritability of spot count by the selection differential to predict the response to selection for your experiment.

5. Run your selection experiment and compare your predicted response to the actual response

Run your simulation for 200 days. Paste a screenshot of your histogram showing the distribution of spot counts among the offspring of your selected parents (this is your current Right Field population). Record the average number of spots among the offspring of the selected parents. Calculate and record the actual response to selection for your experiment.

6. How accurate was your prediction? If it was not accurate, try to figure out why.

7. What was the realized heritability in your experiment?

Triturus balanced lethal system SNP practicum

Willem Meilink, Anagnostis Theodoropoulos, James France & Ben Wielstra

In this practicum you will learn about an extreme hereditary disease known as a balanced lethal system that has independently evolved in some newts, plants and insects. You will also learn about a technique called KASP genotyping that allows you to genotype dozens of markers quickly and economically. This part of the course starts with DNA extraction of newt embryos. The week after the balanced lethal system and genotyping technique are introduced in two lectures. The lectures are followed by a practicum in which the genetic dataset collected for the extracted embryos is analyzed. Altogether there are **42** questions to be answered during the two lectures and the practicum. Answers need to be compiled in a single document that needs to be uploaded as a PDF file in Brightspace. Students work in couples. However, note that **every student needs to upload their own file with answers**.

Introduction

The evolutionary mystery of balanced lethal system

Natural selection is supposed to maximize individual fitness (Darwin, 1859). In this context, **Balanced lethal systems** represent an evolutionary enigma, because they kill off exactly half of the offspring during embryonic development (Grossen et al., 2012; Muller, 1918; Wielstra, 2020). They work as follows: in a diploid organism, there are two forms of a particular chromosome, let us call them *A* and *B*. These two chromosome forms do not recombine (they do not exchange genetic material between each other, rather they are inherited as if they were a single unit, a 'supergene'). Therefore, *A* and *B* are maintained over the generations. Both forms are required for survival and therefore all adults in the population will possess one copy of form *A* and one copy of form *B*. These adults randomly transmit either form *A* or form *B* in each of their haploid egg or sperm cells. Upon fertilization, two gametes fuse to form a diploid zygote. Four combinations are possible, each with a 25% chance, following the rules of Mendelian inheritance: *AA*, *AB*, *BA* or *BB*. Only two of these options, the *AB* and *BA* heterozygotes, are viable; the other half, the *AA* and *BB* homozygotes, perish.

Why would the *AA* and *BB* homozygotes in a balanced lethal system be doomed and express arrested development? It is assumed that form *A* lacks crucial genes that are only present on form *B*, and the other way around (Meilink et al., 2021). Such a missing, dysfunctional gene copy represents a 'lethal allele'. It follows that such key genes involved in a balanced lethal system either have zero or two copies in both types of homozygote. On the other hand, in viable heterozygotes, key genes are expected to be present in a 'hemizygous state', i.e. as a single copy. It is important to remember that only these viable heterozygotes could potentially contribute to the next generation. In contrast, 'normal' diploid genes positioned on the rest of the genome and not involved in the balanced lethal system will show regular Mendelian inheritance; each individual will possess two copies, one inherited from each parent. Hemizygous inheritance being at the basis of balanced lethal systems is purely theoretical at this stage. The research questions we will address are:

- ***Do the genes involved in a balanced lethal system show hemizygous inheritance?***
- ***Do the collection of genes on A and on B behave as supergenes?***

Triturus newts as a textbook example

To address our research questions, we work with the most famous and best-studied balanced lethal system, named 'chromosome 1 syndrome'. For this balanced lethal system, both the evolutionary clade affected (the monophyletic genus *Triturus*, comprising the crested and marbled newts) and the genomic region responsible (the long arm of chromosome 1), are clearly delineated (Grossen et al., 2012; Macgregor and Horner, 1980; Wielstra, 2019). From karyotyping it is known that all adult *Triturus* newts consistently carry two forms of chromosome 1, called *1A* and *1B*. The 50% of their offspring that is homozygous, either *1A1A* or *1B1B*, dies at the late tail-bud stage, approximately halfway through embryological development. Hence, half of *Triturus* eggs never hatch!

To test if inheritance is hemizygous in chromosome 1 syndrome we apply the following rationale. There are three embryo classes that differ in the number of alleles they possess for markers linked to *1A* or *1B*. A *1A1A* homozygote possesses two alleles for a *1A*-linked marker and zero alleles for a *1B*-linked marker, while the opposite applies to a *1B1B* homozygote. On the other hand, a heterozygote (*1A1B* = *1B1A*) possesses one *1A*-linked marker and one *1B*-linked marker. If two species have species-diagnostic and **individually recognizable alleles** for particular *1A* or *1B*-linked markers, the number of alleles present (0, 1, or 2) can be genotyped in their F1 hybrids – they can be counted. For regular, diploid markers, two distinct alleles should always be present.

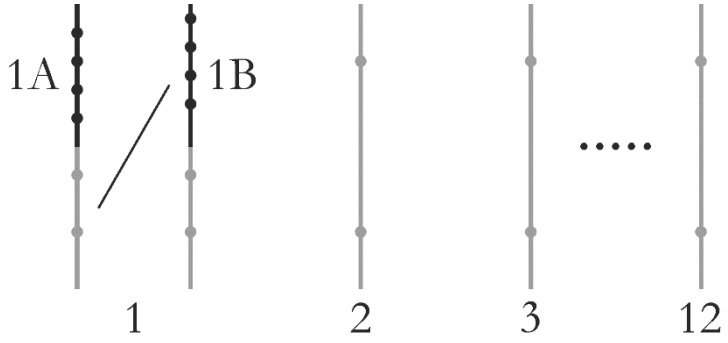
To test if *1A* and *1B* are inherited as supergenes we apply the following rationale. When F1s are crossed with each other to produce F2s, recombination should start shuffling the genomes of the two parental species. However, this would not be the case for genomic regions that are shielded from crossing over during recombination. Such regions (e.g. *1A* or *1B*) would be expected to be transmitted to the F2 generation as a single unit.

Samples, marker design and KASP genotyping

We use hybrids between the Balkan crested newt (*Triturus ivanbureschi*) and the Macedonian crested newt (*T. macedonicus*). These two species are distributed in south-eastern Europe and Turkey and they hybridize in the wild where their ranges meet (Wielstra et al., 2017). Colleagues from the University of Belgrade in Serbia have established a breeding colony of both species and have produced hybrid embryos for us.

To test hemizygous inheritance in the *Triturus* balanced lethal system genetic markers are needed. We have mined a dataset of several thousand of nuDNA markers that has been sequenced for many *Triturus* newts across the natural range (Wielstra et al., 2019; unpublished data). For these markers sequence data are available and their position on the genome is known (linkage map produced by James France, unpublished data).

We used this dataset to select 32 species-diagnostic single nuclear polymorphisms (SNPs) for *T. ivanbureschi* versus *T. macedonicus* that are distributed all across the genome (see Figure below). There are four markers on 1A, four markers on 1B, two markers on the recombining part of chromosome 1 and two markers on each of the remaining chromosomes 2 to 12. We can genotype these 32 SNP markers for each of our embryos, to determine for each SNP marker if *T. ivanbureschi* and/or *T. macedonicus* alleles are present.



KASP (Kompetitive Allele-Specific PCR) is a PCR-based technique that involves fluorescence-based genotyping (Semagn et al., 2014). In a run of KASP three primers are used: one common reverse primer, and two forward primers which only differ at the last nucleotide, complementary to the two possible SNPs. The specific forward primers are linked to a specific fluorescent label. If the SNP is present in the target DNA, the respective forward primer(s) will bind. After multiple rounds of PCR, the fluorescent signal will intensify. The presence of zero, one or two SNPs will result in no, one or two fluorescent signals. KASP genotyping is implemented at IBL's SNP Genotyping Facility.

References

Darwin, C., 1859. On the origin of species. John Murray, London.

Grossen, C., Neuenschwander, S., Perrin, N., 2012. The balanced lethal system of crested newts: a ghost of sex chromosomes past? Am. Nat. 180, E174-E183.

Macgregor, H.C., Horner, H., 1980. Heteromorphism for chromosome 1, a requirement for normal development in crested newts. Chromosoma 76, 111-122.

Meilink, W., France, J., de Visser, M., Wielstra, B., 2021. Balanced lethal systems: an evolutionary mystery. Frontiers for Young Minds 9, 632945.

Muller, H.J., 1918. Genetic variability, twin hybrids and constant hybrids, in a case of balanced lethal factors. Genetics 3, 422-499.

Semagn, K., Babu, R., Hearne, S., Olsen, M., 2014. Single nucleotide polymorphism genotyping using Kompetitive Allele Specific PCR (KASP): overview of the technology and its application in crop improvement. Mol. Breed. 33, 1-14.

Wielstra, B., 2019. *Triturus* newts. Curr. Biol. 29, R110-R111.

Wielstra, B., 2020. Balanced lethal systems. Curr. Biol. 30, R742-R743.

Wielstra, B., Burke, T., Butlin, R.K., Arntzen, J.W., 2017. A signature of dynamic biogeography: enclaves indicate past species replacement. Proc. Royal Soc. B 284, 20172014.

Wielstra, B., McCartney-Melstad, E., Arntzen, J.W., Butlin, R.K., Shaffer, H.B., 2019. Phylogenomics of the adaptive radiation of *Triturus* newts supports gradual ecological niche expansion towards an incrementally aquatic lifestyle. Mol. Phylogenet. Evol. 133, 120-127.

DNA extraction

DNA extraction is a rather laborious procedure. However, keep in mind that it is a fundamental step in most molecular biology work. The success of subsequent experiments is completely dependent on the quality and quantity of the DNA samples!

Read each step of the DNA extraction protocol carefully and continuously keep in mind where the DNA is, so that you do not throw away the wrong tube/solution and lose your DNA.

One of the biggest problems of handling DNA extracts is degradation caused by DNases. DNases are present on our skin. For that reason, when you handle DNA extracts you always wear gloves to protect your sample.

Samples

You will work in couples and extract a single embryo per couple. Each embryo has a unique code, remember the code of your embryo and **make sure you label your tubes with that code, followed by your group number, both on the cap and on the side of the tube!**

Embryos belong to six classes in equal ratio:

- F1 hybrids collected before the embryonic stage in which developmental arrest occurs
- F1 hybrids that have experienced developmental arrested
- F1 hybrids that have survived
- F2 hybrids collected before the embryonic stage in which developmental arrest occurs
- F2 hybrids that have experienced developmental arrested
- F2 hybrids that have survived

You do not know in advance which class your embryo belongs to, it is up to you to figure that out once you get the KASP genotyping dataset!

DNA extraction protocol

Before you start, confirm that you have a heating block set to 55°C (because it takes time to warm up). Eppies with the chemicals required will be handed to you by the organizers of the practicum. Note that proteinase K and isopropanol need to be kept on ice.

Cell Lysis

1. Label a 1.5 ml microcentrifuge tube with the number of the embryo you were assigned and your group number.
2. Pick up the embryo with your tweezers and place it in a new 1.5 ml microcentrifuge tube.
3. Dispense 300 µl of Nuclei Lysis Solution in the new 1.5 ml microcentrifuge tube.
4. Get the Proteinase K (20 mg/ml) from the freezer.
5. Dispense 6.0 µl Proteinase K (20 mg/ml) in the 1.5 ml microcentrifuge tube.
6. Crush the egg with a pestil.
7. Mix with a brief vortex (5-10 seconds).
8. Incubate at 55 °C until tissue has mostly digested (no solid material left). This takes approximately 1-1,5 hours. In some cases the egg shell is not completely digested. After the time elapses show your tubes to a teaching assistant to confirm proper digestion.
9. Vortex every 15 minutes (5-10 seconds).
10. If after 1 hour the tissue is still intact, add another 3 µl of Proteinase K and continue incubation.
11. When tissue is digested, vortex the tube well (10-15 seconds).

Protein Precipitation

12. Cool lysate to room temperature (until the tube is not hot).
13. Add 100 µl Protein Precipitation Solution to the cell lysate mixture.
14. Vortex vigorously to mix the tube contents until it turns milky white or becomes opaque (20 seconds).
15. Centrifuge at the max speed (10-15 krpm) for 8 minutes.
16. While centrifuging get a new 1.5 ml microcentrifuge tube and name it with number of the embryo you were assigned and your group number.
17. If a solid pellet has formed and there is no floating material proceed to the next step. If the pellet is gelatinous or there is still floating material in the supernatant more centrifugation may be required – up to 15 minutes for some samples.
18. Carefully pipette out the supernatant (which contains the DNA) into the new 1.5 ml microcentrifuge tube. It is important to avoid transferring any solids, so try to keep the tip of the pipette just under the surface of the solution, and draw up the liquid slowly. It is better to leave some liquid behind than to transfer solid matter. If the pellet is disturbed during pipetting repeat centrifuge step.

DNA Precipitation

19. Get 100% cold isopropanol from the freezer.
20. Add 300 µl 100% cold isopropanol.
21. Gently invert the tube 50 times to thoroughly mix the contents. In between inversions, take a moment to observe: you might spot the precipitated DNA appearing as white strands.
22. Centrifuge at the max speed (10-15 krpm) for 5 minutes.
23. Carefully remove/poor off the supernatant (precipitated DNA should be visible as white pellet). *Caution! Pellet may be loose and sometimes it may be necessary to remove the supernatant by pipetting.*
24. Add 300 µl 70% ethanol.
25. Invert the tube 25 times to wash the pellet.
26. Centrifuge at the max speed (10-15 krpm) for 5 minutes.
27. Carefully remove/ Pour off the supernatant. *Caution! Pellet may be slippery and loose so be careful not to pour out pellet.*
28. Any remaining alcohol can be pipetted off the pellet to decrease drying time
29. Put your tube on the side with the lid open on sheet of tissue and air-dry the tubes for 20 minutes. If alcohol droplets persist within the tube, employ a pipette to aspirate and spread them across the surface. Allow an additional 10 minutes for the alcohol to thoroughly evaporate.

DNA Hydration and Storage

30. Add 100 µl of TE buffer.
31. Confirm that the eppie with your DNA extract is properly labelled and hand it in to the organizers of the practicum.

The genotyping of your samples will be done by the organizers of the practicum. Details on the procedure will be provided during a lecture next week and afterwards you will work with the data generated for the *Triturus* embryos. Your DNA extract will be left in a **fridge** overnight (*As DNA needs to resuspend it is important to not freeze the extraction at this stage*). DNA should be stored long-term at -20 or -80 °C.

Questions with lecture on the *Triturus* balanced lethal system

Basics of the balanced lethal system

Triturus suffers from a balanced lethal system. You have learned that balanced lethal systems contain two unique non-recombining forms of a particular chromosome. Two adult *Triturus* each carry one 1A and one 1B chromosome, because both are required to survive. They mate and have offspring.

1. Fill in a Punnett square for chromosome 1A and 1B in this cross

2. Which genotypes are present in the fertilized eggs?

Embryonic arrest will set in at the late tailbud stage during which 50% of all embryos will perish.

3. Which genotypes will survive the balanced lethal system?

4. How many copies of 1A and 1B do such viable embryos possess?

5. Which genotypes are the embryos that will show arrested development?

6. How many copies of 1A and 1B do both classes of non- viable embryos possess?

Assume 1A is p and 1B is q.

7. What are frequencies p and q in a breeding population of Triturus?

8. What will the ratios for 1A1A, 1A1B and 1B1B be in the fertilized eggs according to Hardy-Weinberg equilibrium (p^2 x 2pq x q^2)?

9. What is the allele frequency p and q for embryos that survived the balanced lethal system?

10. What is the genotype frequency p^2, 2pq and q^2 for embryos that survived the balanced lethal system?

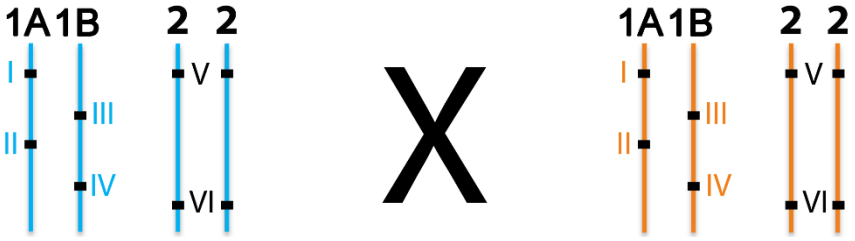
11. What conclusion can you draw about chromosome 1 in Triturus with regards to Hardy-Weinberg equilibrium?

Basics of recombination, or lack thereof

In the genus *Triturus* closely related species can hybridize. For this exercise we assume that hybridization does not negatively affect the fitness of the offspring. You have extracted DNA from embryos from *Triturus*. These were from a cross between *T. macedonicus* x *T. ivanbureschi* (=F1s) or from a cross between two F1 hybrids (=F2s). We will use these to answer the research questions posed in the introduction:

- ***Do the genes involved in a balanced lethal system show hemizygous inheritance?***
- ***Do the collection of genes on A and on B behave as supergenes?***

Below are two hypothetical purebred parents (Ps) used to create the first-generation hybrids (F1s). Chromosomes and alleles are species-diagnostic for one or the other species and are designated with different colors (blue and orange). We look at two genes on chromosome 1A (I and II), two genes on 1B (III and IV) and two genes on chromosome 2 (V and VI).

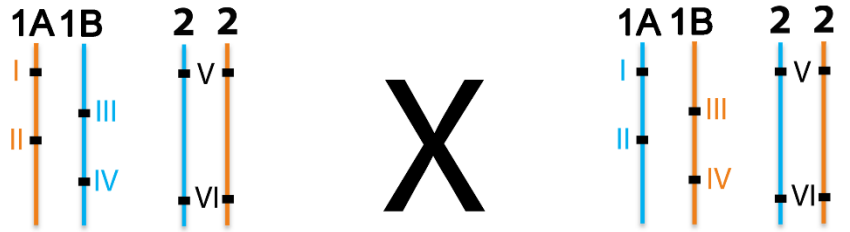


12. Is crossing-over of species-diagnostic alleles possible in the P cross that yields the F1 generation?

13. What would the genotypes of viable F1 offspring look like for chromosome 1?

14. What would the genotypes of viable F1 offspring look like for chromosome 2?

Below are two F1 hybrid parents used to create second generation hybrids (F2s). We can use these hybrids to visualize the prediction that 1A and 1B are inherited as supergenes, whereas for the autosomal 'normal' rest of the genome homologous chromosomes are capable of exchanging sections via crossing over.



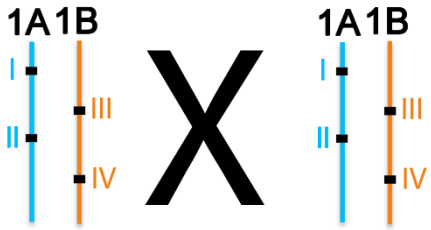
15. Could crossing-over of species-diagnostic alleles occur in F1 individuals (and therefore be observed in the F2 generation)?

16. Does crossing-over occur between two copies of 1A or two copies of 1B and why?

17. What would viable F2 offspring look like for chromosome 1?

18. What would viable F2 offspring look like for chromosome 2?

Below are two other F1 hybrid parents used to create second generation hybrids (F2s):



19. What would viable F2 offspring look like for chromosome 1? Is there a difference with the previous cross (question 17)?

Questions with lecture on KASP genotyping

Consider a species-specific SNP on an autosomal *Triturus* chromosome (excluding 1A or 1B regions). We cross two species.

20. If the genotype of one parent species is A:A and the genotype of the other parent species is G:G, what are the resulting genotypes for the F1 embryos?

21. What would be the expected pattern in the KASP fluorescence graph for the F1 embryos resulting from crossing parents of two different species?

22. Can a parent that is not a hybrid have the genotype A:G ?

Next, consider a species-specific SNP on the (hemizygous) 1A region.

23. If one parent has an allele with nucleotide A and the other parent has an allele with nucleotide G, what are the possible genotypes for the F1 embryos? Explain your answer with a Punnet's square.

24. What would be the expected pattern in the KASP fluorescence graph for the F1 hybrids resulting from crossing two different species? Based on the Punnett square, can you assign the genotypes of the F1 embryos to the image?

Consider we cross two F1 hybrids that both have genotype A:G for a SNP in an autosomal chromosome (so excluding the 1A and 1B region) to produce F2 hybrids.

25. What are the possible genotypes in the F2 generation? Explain your answer with a Punnet's square.

26. What would be the expected pattern in the KASP fluorescence graph for the F2 embryos? Based on the Punnett square, can you assign the genotypes of the F2 embryos to the image?

Consider a SNP in the 1A region where one species has nucleotide A and the other species has nucleotide G.

27. If we cross two F1 hybrid individuals, what are the possible genotypes for the F2 embryos? HINT: The F1 individuals must be viable to reproduce.

28. Is it possible for an F2 embryo to inherit two 1A chromosomes from the same species?

Questions based on working out the SNP genotyping data

Hemizygous inheritance and selection in the balanced lethal system

Download and inspect the data file

Download the dataset "KASP_practicum.xlsx" with the genotype data for the individuals you extracted last week from Brightspace and open it in Microsoft Excel or a similar spreadsheet program. Rows represent individuals. On top you see genotypes of the parents of the F1 and F2 embryos. Below are the genotypes of all the embryos for which DNA was extracted last week, including yours. Note that embryos are allocated to six groups, labelled 1-6 in the first column, that correspond to the six embryo classes that have been genotyped (see Samples, marker design and KASP genotyping). It is up to you to figure out which embryos belong to which class. Columns C-AH represent the 32 nuclear DNA SNP markers genotyped. Note that these simply have a number, it is up to you to figure out which markers are positioned on 1A or 1B or on the rest of the genome (see Samples, marker design and KASP genotyping).

Text to columns transformation

Before you can work with your data, you need to reformat the file. To conduct calculations later, we need to split the genotype call for each marker over two columns, using the "Text to Columns" function. **Be careful not to copy over information in adjacent columns.**

1. Open the " KASP_practicum.xlsx" File in Excel on the "PART 1" sheet
2. We want to insert additional empty columns on the right side of each genotype call (i.e. columns C – AH). Hold down the "Ctrl" key and left-click on the top of each column (one by one) in the column bar, starting from column D (not C) and ending with column AH. Each column should be highlighted when selected.
3. Right-click anywhere on the selected columns, and then left-click on "Insert." This will insert an empty column on the left side of each of the selected columns and creates space for further analysis.
4. Left-click on the top of the column in the column bar that corresponds to the first marker, which is column C.
5. Go to the "Data" tab on the Menu Bar and left-click on the "Text to Columns" button that appears on the Ribbon bar. This action will open the "Convert Text to Columns Wizard" window.
6. In the first step, select the option "Delimited," and then click on the "Next" button to proceed.
7. Under "Delimiters," click on "Other" and input a colon (":") in the corresponding text box. Then click on the "Finish" button to apply the transformation.
8. If a pop-up window appears, asking if you want to replace existing data, click "OK" to proceed. Note that this transformation will not affect the original data but will split it into multiple columns based on the chosen delimiter (colon).
9. For each of the markers (from column C to AH), repeat the process (steps 4-8) one column at a time. This step is crucial to appropriately separate the relevant information and prepare the dataset for further analysis.

Conditional formatting

Next, we will use conditional formatting to color SNP alleles according to parental species. This allows you to visualize which species-diagnostic alleles have been passed down the generations. We will use two different colors for the SNP allele representing each of the two parental species for each marker (orange for *T. ivanbureschi* alleles and blue for *T. macedonicus* alleles). Missing data is noted as "?". If you want to give this a distinct color you need to specify this symbol as "~?", because Excel otherwise interprets a "?" to represent any character. But note that a cell with a "?" will remain blank and therefore would stand out if all other cells are colored.

1. Left-click on the top of column C in the column bar, hold the click and drag the cursor to the top of column D. Both columns C and D (i.e. the first marker) are now highlighted.
2. Go to the "Home" tab on the Menu Bar and left-click on the "Conditional Formatting" button that appears on the Ribbon bar.
3. Under "Highlight Cell Rules," select "Text That Contains…". This action will open the "Text that Contains" window.
4. In the text box, write down the nucleotide that corresponds to the *T. ivanbureschi* parent for that marker. For example, if the nucleotide is "T," type "T" in the text box.
5. Click on the drop-down menu on the right side and choose "Custom Format…".
6. On the "Format cells" window click on the Fill tab and choose an orange color then click "OK." All cells containing the specified nucleotide "T" in the selected columns (C and D) will now appear with the chosen formatting.
7. With the two columns (C and D) selected, repeat steps 2 and 3, but this time write down the nucleotide that corresponds to the *T. macedonicus* parent for that marker. For instance, if the nucleotide is "C," type "C" in the text box.
8. Click on the drop-down menu on the right side and follow step 6, but this time choose a blue color. Then click "OK." All cells containing the specified nucleotide "C" in the selected columns (C and D) will now appear with the chosen formatting.
9. Repeat steps 1 to 7 for all the markers in the dataset. **Remember to change the nucleotides you format for each marker, but maintain consistent colors (orange for *T. ivanbureschi* and blue for *T. macedonicus*) throughout the entire dataset.**

29. What does missing data mean for a (normal) diploid marker? What does missing data mean for a hemizygous marker?

30. The genotype for the T. ivanbureschi parent for marker 10 is missing. Based on the results for the other individuals, what is the genotype of the T. ivanbureschi parent?

31. Which markers are positioned on either A or B and why?

32. Which three embryo groups are the F1s and which three groups are the F2s and why?

33. Which two groups of embryos survived the balanced lethal system and why?

34. Which two groups have perished from the balanced lethal system and why?

35. Which two groups of embryos have been collected before developmental arrest in the balanced lethal system took place and why?

36. What is your sample number and to what embryo class does it belong?

37. Marker 9 and 10 are positioned on the same chromosome. Can you give an example of an instance of crossing over?

Linkage and recombination

Inspect the data file

Go to the “PART 2” sheet of the "KASP_practicum.xlsx" file. In this sheet we provide a dataset that consist of 4 markers for 75 individuals (including two purebred parents used to breed F1 individuals and four F1 hybrid parents used to breed F2 individuals). Markers 1 and 2 are located on chromosome 1A and markers 3 and 4 are located on chromosome 5. The F2 offspring are all arrested individuals that possess two 1A chromosomes. The main objective of this exercise is to manually calculate the linkage between these markers and interpret your results.

Principles of linkage and haplotypes

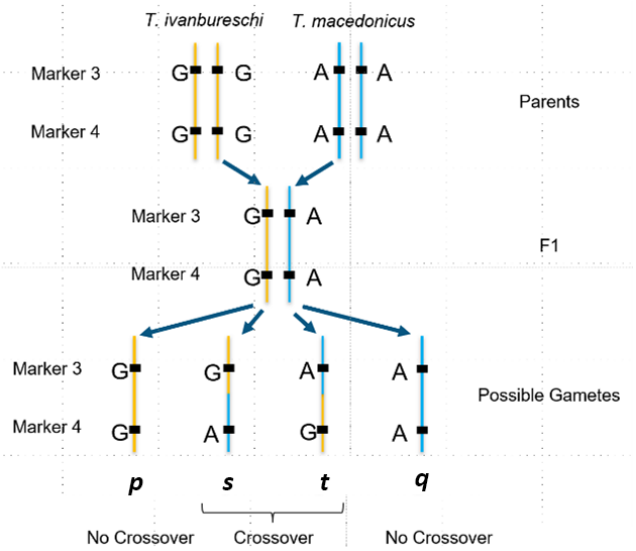
Calculating linkage will allow us to assess the co-inheritance pattern of linked markers and detect potential crossovers that occurred during meiosis. It will also allow us to assess if two markers are in linkage equilibrium, meaning that they are inherited randomly (i.e. they are not linked) or if they are in linkage disequilibrium (often abbreviated as LD), where there is a non-random association of alleles (i.e. they are linked). Linkage disequilibrium is not a yes/no situation: the further apart two markers on the same chromosome are, the larger the chance that a crossing over event between them occurs.

Calculating linkage involves transforming the genotypic information into haplotypes. A **genotype** refers to the alleles or variants an individual carries in a particular genetic locus, while a **haplotype** refers to a group of alleles in an organism that are inherited together from a single parent. For instance, let us look at the two linked markers 3 and 4 and their respective genotypes. Both markers 3 and 4 have the genotypes: G:G for *T. ivanbureschi* and A:A for *T. macedonicus*. To determine the haplotypes, let us look at the chromosomes of the parents and the F1 hybrids. The *T. ivanbureschi* parent has a “GG” haplotype, while the *T. macedonicus* parent has a “AA” haplotype. The resulting F1 hybrids will have both haplotypes (“GG” and “AA”). Now lets look at the possible gametes produced by the F1 hybrid parents.

	Chromosome 5			
	3		4	
T.ivanbureschi Parent	G	G	G	G
T. macedonicus Parent	A	A	A	A
F1 Hybrid parent	G	A	G	A
F1 Hybrid parent	G	A	G	A
F1 Hybrid parent	G	A	G	A
F1 Hybrid parent	G	A	G	A

Without crossovers, the gamete haplotypes would be:

- Nonrecombinant chromosome “orange”: “GG”
- Nonrecombinant chromosome “blue”: “AA”



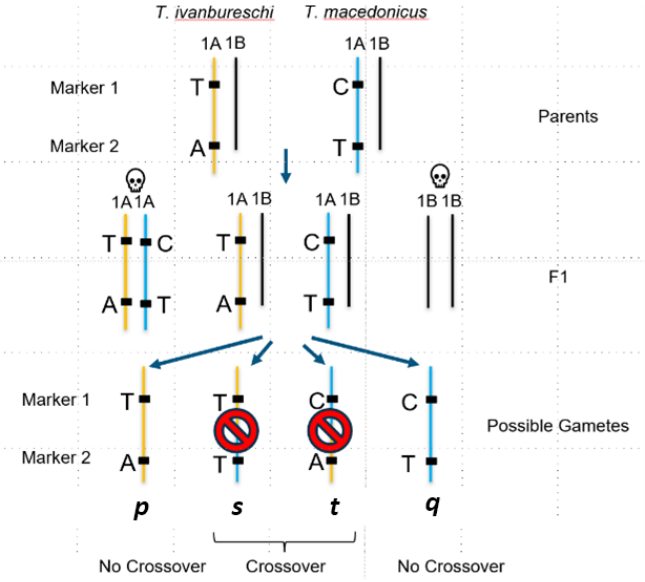
However, in the presence of crossovers, additional haplotypes could arise:

- Crossover event 1: “GA”
- Crossover event 2: “AG”

Transforming the genotypic information into haplotypes is slightly different for the two markers (1 and 2) located on chromosome 1A, given the different (hemizygous) inheritance pattern. Remember that the purebred and F1 parents must have survived the balanced lethal system and therefore only contain a single copy of chromosome 1A (as well as one copy of chromosome 1B). These individuals are not actually homozygotes for markers on chromosome 1A. Remember from the KASP genotyping lecture that these parents only look like a homozygote due to the lack of chromosome 1A markers on chromosome 1B. For example, the *T. ivanbureschi* parent possesses only a single SNP allele for marker 1 and its genotype should be interpreted as T:?. With this insight we can now recreate the inheritance patterns of the markers positioned on chromosome 1A as above. Now lets look at the possible gametes produced by the F1 hybrid parents.

CAUTION: Remember that the (viable) F1 parents only possess a single copy of chromosome 1A. In the figure to the left, one F1 parent has a *T. ivanbureschi* chromosome 1A and the other a *T. macedonicus* 1A (but different combinations are possible).

	Chromosome 1A			
	1		2	
T.ivanbureschi Parent	T	T	A	A
T. macedonicus Parent	C	C	T	T
F1 Hybrid parent	C	C	T	T
F1 Hybrid parent	T	T	A	A
F1 Hybrid parent	C	C	T	T
F1 Hybrid parent	T	T	A	A



Without crossovers, the gamete haplotypes would be:

- Nonrecombinant chromosome "orange": "TA"
- Nonrecombinant chromosome "blue": "CT"

Remember, crossovers do not occur in chromosome 1A, so these additional haplotypes should not arise:

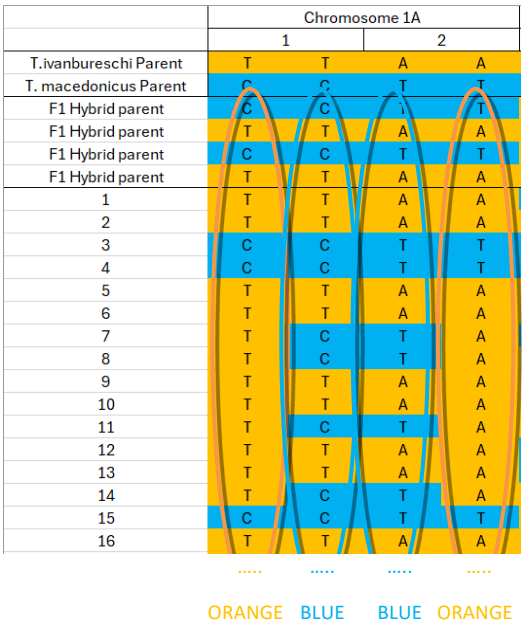
- Crossover event 1: "TT"
- Crossover event 2: "CA"

In this case, we already know that markers 1,2 and 3,4 are not on the same chromosome, but in practice you do not typically have such insight! Therefore, we act in this exercise as if we do not know the relative positions on the genome of markers 1-4: we treat any pair of markers as potentially positioned on the same chromosome. First, we determine the potential "orange" or "blue" haplotypes, based on the parents, and then we follow the inheritance pattern of the potential gametes, for all pairwise combinations of markers (e.g. markers 1 and 2, markers 1 and 3, etc.).

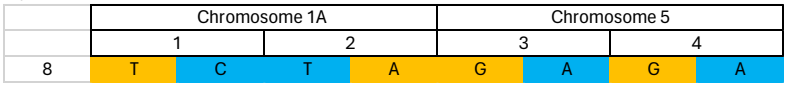
Converting genotypic data into potential haplotypes

1. First define potential haplotypes for all the pairwise marker combinations in the F2s individuals. We start with markers 1 and 2. In cell L8, type "1/2 orange" and in cell M8, type "1/2 blue". Here, "1/2" signifies the pairwise combination of markers, while "orange" or "blue" denotes the columns are used to define the haplotypes. Because the F2s show mixed colors, refer to total data set to identify which column is the "orange" one and which column is the "blue" one. The

column that has the most orange alleles for a marker is going to be the one that corresponds to the “orange” column and the same applies for the “blue” one. *Look at the image bellow to understand the choice of columns.*



CAUTION: *The orange and blue colors in this case are a way for us to not flip around the genotype calls. We use these colors to keep allele order consistent during genotype calling. To prevent flipped genotypes, we always call the "orange" column allele first and the "blue" column allele second. Otherwise, the produced haplotypes we produce would be incorrect. Look at the image bellow for an example of the “orange” and “blue” columns for markers 1 and 2. The individual below has a T:C genotype for marker 1 and a A:T for marker 2, not a T:A like it would appear if we were simply reading all of the genotypes from left to right. Similarly, the individual below has a G:A genotype for marker 3 and a G:A for marker 4, and not A:G.*



2. Create the ”orange” column haplotypes for this particular marker combination: In cell L9 enter the formula: =Cell1&Cell2. For Cell1, use the first cell in marker 1 that holds a nucleotide for the "orange" column. For Cell2, utilize the first cell in marker 2 that holds a nucleotide for the "orange" column.
3. Autofill the haplotypes for this particular marker combination: Select cell L9. At the bottom-right corner of the cell, find the small square. Left-click on it and drag your cursor down to cell L77 to autofill. This action fills in the haplotypes for the "orange" column for all F2s.
4. Repeat the process for the "blue" column: Begin with cell M9 and follow steps 2 and 3.
5. Now repeat the above steps for the remaining marker combinations. **Do not forget to change the cells!**

By following these steps, you have transformed the genotypic data into potential haplotypes for all pairwise combinations of markers. The next step involves calculating their frequencies by counting how often each potential haplotype occurs and then dividing this number by the total number of haplotypes called (which is two times the number of F2 individuals).

Calculating Haplotype Frequencies and D

Linkage disequilibrium (LD) is commonly estimated using *D*, which quantifies the deviation of haplotype frequencies from the expected values based on allele frequencies. The formula for calculating *D* is:

$$\boldsymbol{D = pq - st}$$

In this formula, *p* and *q* are the frequency of the potential haplotypes when no recombination occurred, while *s* and *t* are the frequency of the potential haplotypes produced by a crossing over event (see figures above).

The values of *D* range from -0.25 to 0.25, where the maximum and minimum values indicate absolute linkage, while *D* = 0 suggests that the loci are in linkage equilibrium, which means that no linkage exists between them.

1. Generate four cells labeled with the letters *q*, *p*, *s*, and *t* below the potential haplotypes in cells K80 to K83. These letters correspond to the frequencies of the "TA", "CT", "TT", and "CA" haplotypes, respectively, for the 1/2 marker combination.
2. Calculate the frequency of the "TA" haplotype in cell L80, using this formula: =COUNTIF(L9:M77; "TA")/COUNTA(L9:M77) (double check if the range of cells provided here is correct for your dataset).
3. Repeat the previous step for the other possible haplotypes in cells L81 to L83. Keep the range constant while altering the formula to correspond to each specific haplotype of interest.
4. Input "*D*" in cell K85, and calculate *D* in cell L85, calculate *D* using this formula: =L80*L81-L82*L83.
5. Replicate this entire process for all other marker combinations. **Remember to change the ranges and haplotypes according to the particular marker combination**.

By following these steps diligently for each marker combination, you will successfully calculate the frequencies of the potential haplotypes as well as *D*.

38. Of the potential haplotypes from markers 1 and 2, which ones are absent in the dataset and why?

39. What are the corresponding potential haplotypes for markers 3 and 4 to the letters p, q, s, and t?

40.What are the corresponding potential haplotypes for markers 1 and 4 to the letters p, q, s, and t?

41. Fill in a cross table with the D values that you calculated and explain the results. What do these values say about the linkage of the markers?

	1	2	3	4
1	-			
2		-		
3			-	
4				-

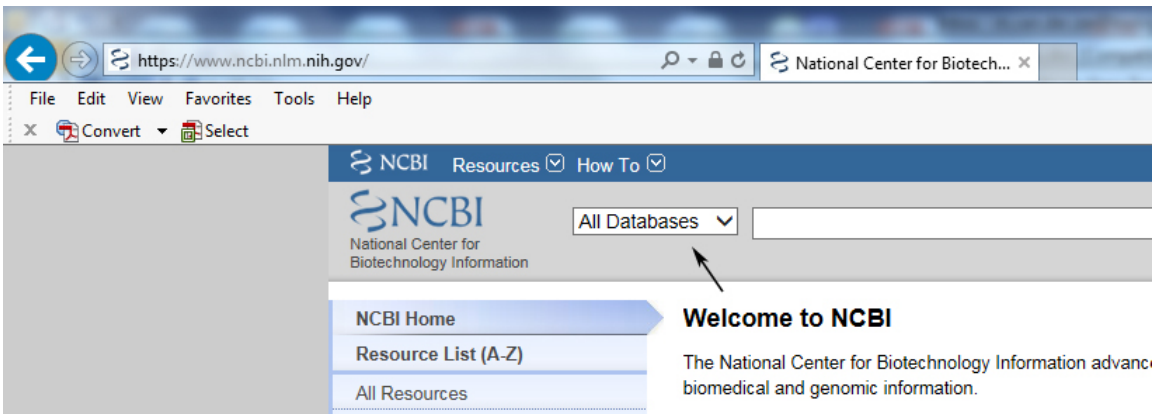
42. What is the calculated D value for markers 1 and 2? What does this value mean regarding linkage?

Computerpracticum Gene Trees

In dit computerpracticum gaan we proberen te achterhalen wanneer de duplicatie is opgetreden van twee verwante genen die we in *Drosophila melanogaster* vinden, namelijk *tolloid* en *tolkin*. Antwoorden op de vragen moeten in een enkel PDF bestand verzameld worden en **geüpload worden in Brightspace door elke individuele student**.

1. Zoek eerst de eiwitsequentie van Tolkin op.

Dit kan op vele manieren. Bijvoorbeeld op NCBI http://www.ncbi.nlm.nih.gov. Je kunt daar direct in het venster "All databases"(zie pijltje in screenshot) aangeven dat je een eiwit zoekt en bijvoorbeeld "*Drosophila melanogaster* tolkin" intypen.

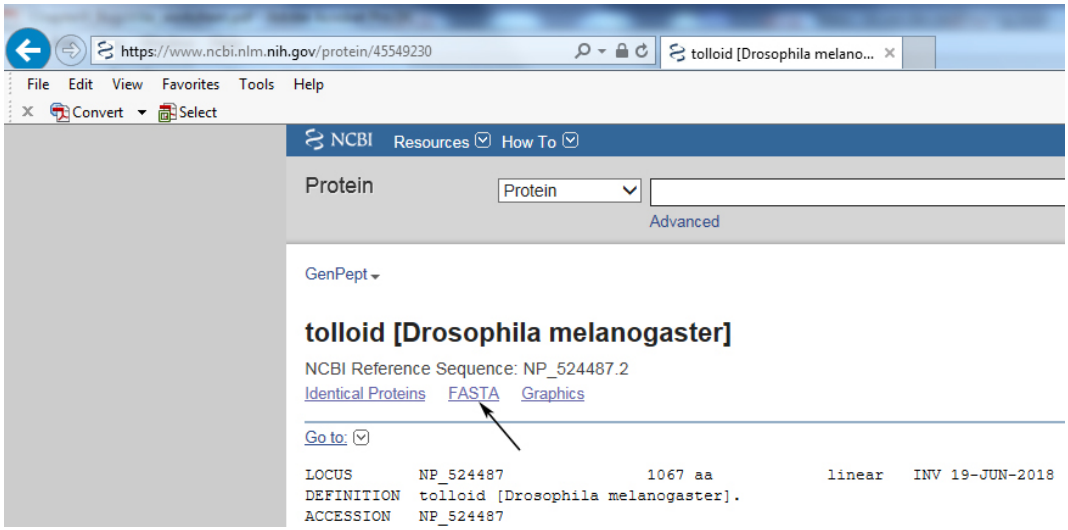


Je zult zien dat er vele sequenties beschikbaar zijn. Je ziet vaak het woord "isoform".

A. **Wat betekent "isoform" precies? Maakt het voor ons uit wat voor isoform je neemt?**

Onder elke entry staat hoeveel aminozuren lang de sequentie is (aa). Scroll naar beneden; daar zie je ook kortere sequenties (wat is daar aan de hand denk je?). Klik naar een volledige eiwitsequentie.

Als je in die entry helemaal naar beneden scrolt, zie je de eigenlijke sequentie. Die sequentie willen we hebben in fasta format. Om je te helpen, kun je helemaal bovenaan de pagina de link "FASTA" klikken (zie pijltje in screen shot hieronder)



Een fasta file heeft het volgende format:

```
>Naamvanjegen
SEQUENTIE
```

Copy-paste de sequentie, in het fasta format, naar een text editor, Notepad bijvoorbeeld. Uiteindelijk gaan we één grote FASTA file maken met alle sequenties die we vinden.

2. We gaan nu kijken of er verwante sequenties binnen *Drosophila melanogaster* te vinden zijn, die door een genduplicatie zouden kunnen zijn ontstaan. Dit kun je doen met BLAST, bijvoorbeeld op NCBI: http://blast.ncbi.nlm.nih.gov/Blast.cgi (kies protein blast). Copy-paste je sequentie in het Query Sequence veld. In het veld "Organism" kun je aangeven dat je alleen sequenties in *Drosophila melanogaster* wilt zoeken. Verander verder geen instellingen en ga zoeken.

Onder het tabje "Descriptions", zie een lijst van sequenties die op jouw zoeksequentie lijken.

B. **Wat geven de waardes "Query cover", "E value" en "Ident" weer?**

Wat zie je aan de waardes na de derde hit? Kijk ook eens onder het tabje "alignments" Wat is er denk je aan de hand met de eerste 3 hits? Uiteindelijk zijn we op zoek naar de sequentie van een ander gen dat door duplicatie zou kunnen zijn ontstaan, niet naar de sequentie van Tolkin zelf (dus niet naar 99 of 100% identity).

Kijk nog eens naar de waardes onder het tabje "Descriptions". Wat is er denk je aan de hand met de 6e hit? (vergelijk met 5!). En vanaf hit 8? Wat is daar aan de hand? (kijk naar de namen van de sequenties). Wanneer worden de hits onzin denk je? (kijk ook eens naar de alignments voor die hits). Hit 7 is lastig; wat denk je daarvan?

C. **Alles overwegend, hoeveel paralogen (andere genduplicaten) van Tolkin zijn er nu aanwezig in *Drosophila melanogaster*?**

Kies uiteindelijk de eiwitsequentie van Tolloid. Copy-paste die sequentie, weer in fasta-format, onder de sequentie van Tolkin in dezelfde tekst editor file. Dat ziet er dan zo uit:

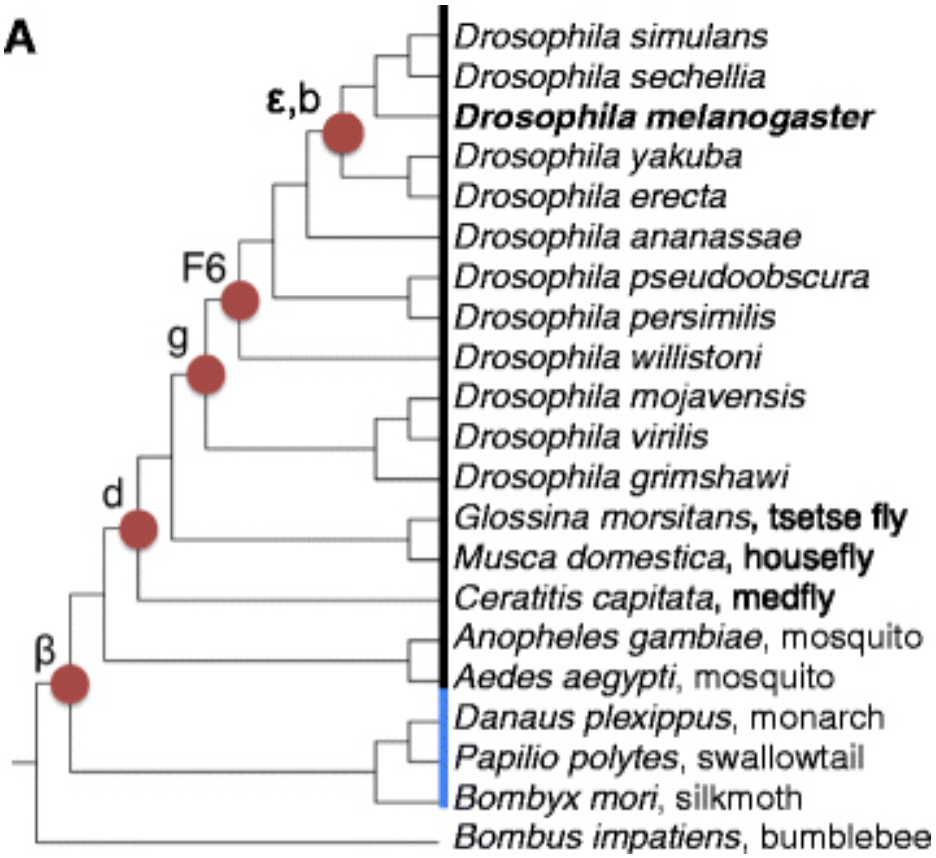
```
>Sequentienaam1
SEQUENTIE

>Sequentienaam2
ANDERESEQUENTIE
```

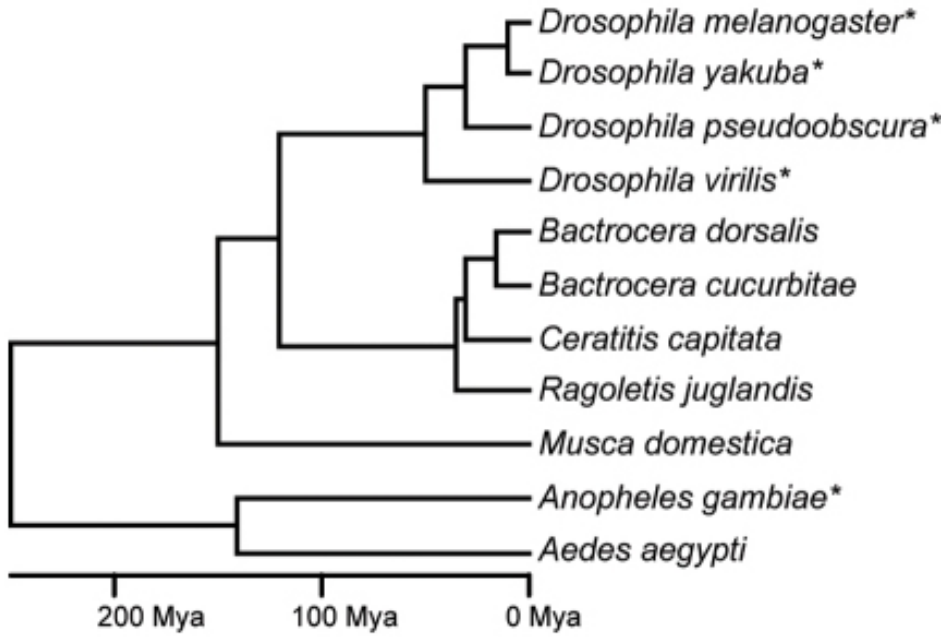
Uiteindelijk gaan we één hele grote FASTA file maken met alle sequenties, ook uit andere organismen, waarvan we een gene tree gaan maken. Het is handig de sequenties duidelijke namen te geven, zodat je zelf nog weet wat wat was. Voor de fylogenie-programmaatjes is ook handig als de eerste 8 letters van elke naam al meteen verschillend zijn." >DmelTolkin" zou bijvoorbeeld een goede naam zijn voor *Drosophila melanogaster* Tolkin.

De twee sequenties die je nu hebt, zijn ooit eens door een genduplicatie ontstaan. Uiteindelijk willen we achterhalen wanneer deze duplicatie heeft plaatsgevonden. Pas in de fruitvliegen? In een voorouder van alle insecten? Al in een voorouder van alle arthropoden? Alle dieren?

3. We gaan nu proberen in andere organismen orthologen van onze genen te vinden. Op de volgende pagina staan een paar stambomen om je te helpen bij het kiezen van andere soorten.



(Sawyer et al., 2017)



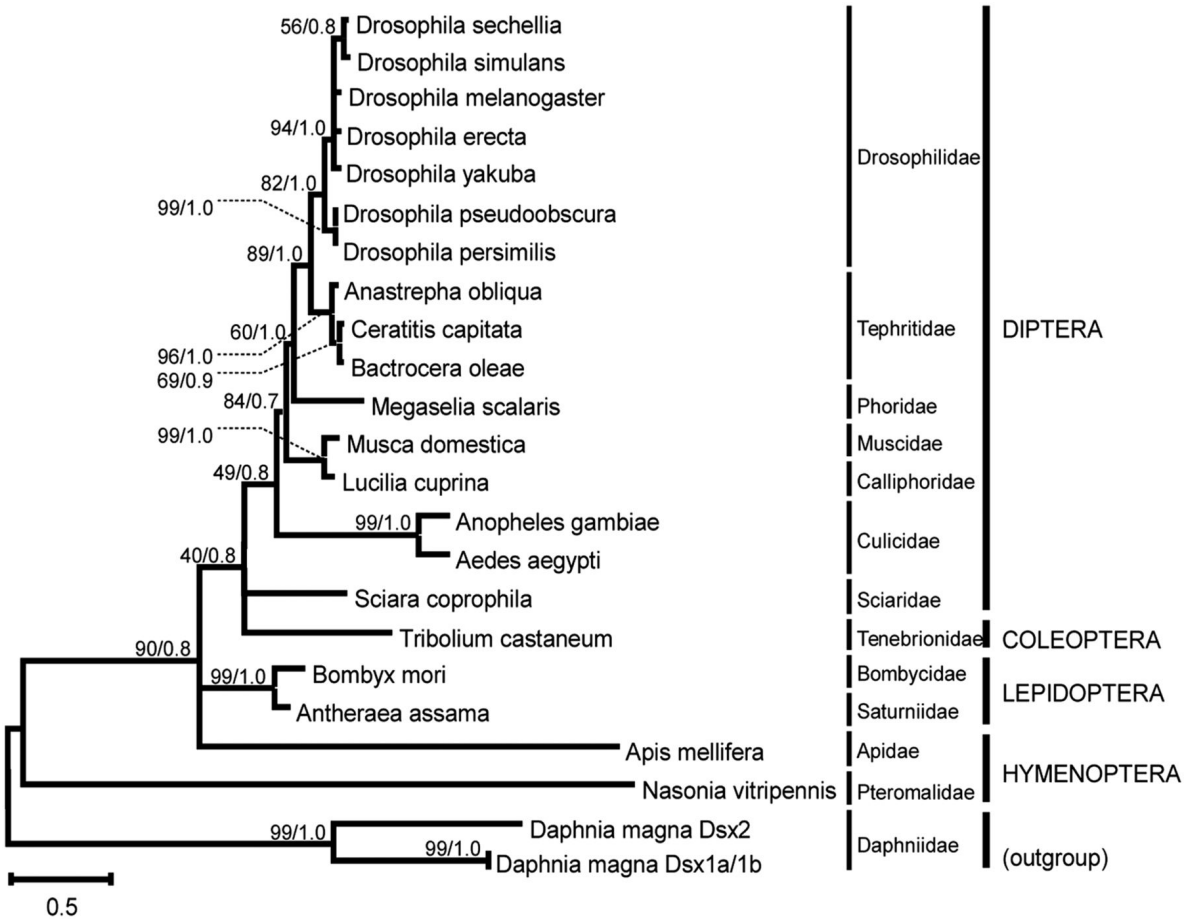
(Petersen et al., 2009)

We gaan eerst eens kijken of we beide genen vinden binnen het geslacht *Drosophila*, bijvoorbeeld in *Drosophila pseudoobscura*. Ga weer naar de protein blast site, geef aan dat je sequenties van *Drosophila pseudoobscura* zoekt.

Blast Tolkin. Kijk eens naar de verdeling van Per Ident. waardes van de eerste 4 hits. Blast Tolloid en kijk ook naar de verdeling van de Per Ident waardes van de eerste 4 hits. Wat is er aan de hand denk je? Kijk onder het tabje “alignments” ook eens naar de verdeling van de lengtes van de gevonden eiwitten. De namen van de sequenties zijn door een computer d.m.v. automatische annotaties toegewezen en slaan soms nergens op. Hoeveel paralogen zijn er nu denk je in *Drosophila pseudoobscura*?

Kies zowel voor Tolloid als voor Tolkin de orthologe sequentie in *Drosophila pseudoobscura*, geef ze de juiste naam en voeg ze toe aan je FASTA file. Kijk nu of je beide *melanogaster* sequenties nog verder terug kunt vinden in de fruitvliegen-fylogenie, bijvoorbeeld in *Drosophila virilis*. Beoordeel de verschillende hits op dezelfde manier. Zet ook de *virilis* sequenties in je FASTA file.

We zullen dus buiten *Drosophila* moeten kijken om deze genduplicatie te plaatsen. Hoe zit het in de Middellandse zeevlieg (*Ceratitis capitata*). En in de huisvlieg (*Musca domestica*)? Zet de gevonden sequenties in je FASTA file.



Ga zo verder. Hoe zit het in de muggen? (bijvoorbeeld *Anopheles gambiae).* En in andere insecten? Onderzoek in ieder geval de kever *Tribolium castaneum* (Coleoptera) en de mot *Manduca sexta* (Lepidoptera). Wat is de beste hit voor Tolloid? Wat is de beste hit voor Tolkin? Als beide *Drosophila* duplicaten dezelfde hit geven in één insect, hoef je natuurlijk alleen die ene sequentie in je FASTA file te zetten.

Wat verwacht je voor de Hymenoptera? Zet in ieder geval nog sequentie(s) van de wesp *Nasonia vitripennis* in je file. Je hebt nu een grote FASTA file met alle sequenties die we nodig hebben.

4. We gaan nu een stamboom van de genen maken. Maak daartoe eerst een alignment. Je kunt dit doen met software uit de eerste week, of bijvoorbeeld online met clustal omega op http://www.ebi.ac.uk/Tools/msa/clustalo/. Importeer of copy-paste je fasta file naar de site en klik Submit. Bekijk het alignment Zijn de eiwitten over de hele lengte geconserveerd of slechts in domeinen? Bij "Phylogenetic tree" kun je de stamboom bekijken.

D. **Wat is dit nou voor boom? Is het een echte fylogenie, i.e. een cladogram? Waarom wel/niet? Wat is er uiteindelijk voor methode gebruikt om de sequenties te rangschikken?**

Je kunt de stamboom in newick format opslaan (dat format met haakjes). Als je die info copy-paste in een text file , heb je alles wat je nodig hebt. De software MEGA is superhandig om de stamboom mee te editen. Dowload MEGA van https://www.megasoftware.net/ en installeer het op je computer. Onder "User trees" kun je met "Display newick trees" de stamboom bekijken (wellicht even show "all files" aanklikken als je je opgeslagen boom niet kunt vinden). Je kunt nu met de buttons links de boom ergens rooten. Je kunt ook takken draaien, zodat de organismen in een logische volgorde komen te staan.

E. **Wat is het beste taxon om de boom te rooten?**

F. **Komt je stamboom van de eiwitten overeen met de daadwerkelijke stamboom van de diersoorten? Waar wel/niet?**

G. **Je oorspronkelijke twee genen uit *Drosophila* hadden ooit eens een gemeenschappelijke voorouder. Wanneer is die duplicatie ontstaan?**

Je moet van elke mogelijke combinatie van 2 genen in de stamboom in staat zijn te beantwoorden of deze genen paraloog of ortholoog zijn. Check of dat lukt.

H. **Kijk nog eens naar de branch length bij de aftakking van de Tolloid genen in de vliegen. Wat valt op? Wat vind jij nu de beste naam voor de orthologe genen in de mug, mot, kever en wesp? Je stamboom laat nu goed zien welk gen hoe moet heten. Heb je alle eiwitten de juiste naam gegeven?**

Lever je Gene Tee en de antwoorden op de 8 vragen in.

Practicum: Evo-Devo

RNAi in de meelkever *Tribolium castaneum*

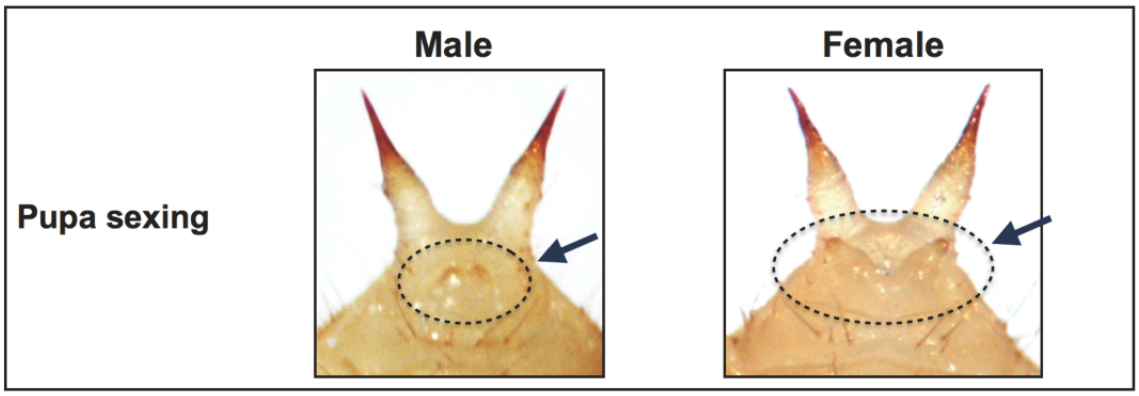
Tijdens dit practicum ga je met RNAi een ontwikkelingsgen uitschakelen in de meelkever *Tribolium castaneum*. Bij RNAi, wordt dubbelstrengs RNA (dsRNA) met de sequentie van een specifiek mRNA geïnjecteerd. Hierdoor wordt dat mRNA gedegradeerd en wordt het eiwit (bijvoorbeeld transcriptiefactor) niet meer gemaakt. Wij injecteren een lang stuk dsRNA. Dat wordt in het insect in kleinere stukjes (siRNAs) geknipt door het enzym DICER. Dan worden de siRNAs geladen op het RISC complex (RNAi Induced Silencing Complex) en enkelstrengs gemaakt, zodat RNA strands ontstaan die complementair zijn aan het mRNA. Het mRNA bindt aan deze complementaire sequenties en wordt door het RISC complex gedegradeerd.

Met behulp van hoofdstuk 11 uit het boek *Evolution* van Barton et al (Brightspace Week 6, 10 Oktober), kun je er naderhand achter komen tegen welk ontwikkelingsgen je RNAi hebt gedaan.

In deze video zie je hoe RNAi praktisch gaat *Tribolium*. Let op, we voeren dus zogenaamde parentale RNAi uit (pRNAi). We injecteren de moeder. Het geïnjecteerde dsRNA komt tijdens de oögenese ook in de eieren terecht, en schakelt daar ook mRNA uit dat zygotisch tot expressie komt. Zo kun je dus door de moeder te injecteren, makkelijk de functie van een ontwikkelingsgen bestuderen in embryos.

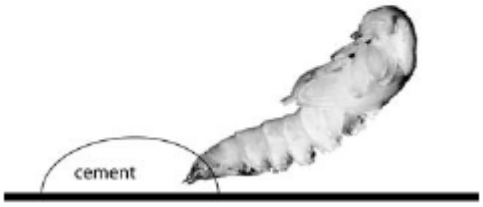
Zeef de poppen uit het meel met de grove zeef (600-850 µm mazen). Pik er onder de stereomicroscoop een stuk of 20 poppen uit met een zachte pincet; het liefst oude poppen die al donker zijn (kijk naar de kleur van ogen, kaken, vleugels, pootjes). Hieronder zie je poppen gesorteerd van jong naar oud:

Scheid de mannetjes en de vrouwtjes. Dat is niet heel moeilijk. Leg de poppen op hun rug. Aan het eind van hun achterlijf zitten twee grote sprieten: de urogomphi. Daar bovenop (ventraler dus) liggen bij de vrouwtjes nog twee kleinere stekeltjes, die je ook van de zijkant kunt zien. De mannetjes hebben dat niet. Bij de mannetjes zie je hoogstens helemaal in het midden soms iets, zie plaatje hieronder:



De vrouwtjes gaan we injecteren. De mannetjes voegen we later ongeïnjecteerd toe. Als je twijfelt of een pop mannelijk of vrouwelijk is, kun je die dus beter bij de vrouwtjes doen; dan wordt die in ieder geval geïnjecteerd. Het zou vervelender zijn als we later met de mannetjes een ongeïnjecteerd vrouwtje toevoegen dat normale eieren kan leggen omdat we geen RNAi hebben gedaan.

Breng, net zoals in het filmpje, een lintje lijm (rubber cement) aan op een objectglaasje en druk heel zachtjes met een zachte pincet het puntje van het achterlijf van de pop een klein beetje op de lijm (niet diep!). Maak een rij van ongeveer 10 vrouwtjes.



Anders dan in het filmpje, proberen we nu met een Hamilton injectiespuit de poppen met een beetje dubbelstrengs RNA te injecteren onder de stereomicroscoop (ongeveer 0,2 µl). Als je klaar bent, kun je het microscoopglaasje zo in een petrischaal met meel zetten, dat de poppen ondersteboven hangen. Als ze uitkomen, vallen de kevers in het meel. Je kunt een kleiner aantal mannelijk poppen er los bij doen. De poppen worden op 30 graden gezet om tot kever te ontwikkelen. In de komende dagen krijg je de kans om te kijken of de kevers zijn uitgekomen en om de objectglaasjes weg te doen. De kevers kunnen dan eieren gaan leggen.

Melkzuurpreparaten maken

We hopen dat de geïnjecteerde vrouwtjes eieren hebben gelegd en zeven deze uit het meel. We zeven daartoe met twee zeven tegelijk: de bovenste, grove (600-850µm) om de kevers uit te zeven en de onderste fijne (250-300 µm) om de eieren op te vangen. Veeg met een penseeltje ook over de bodem van het petrichaaltje om eventueel vast zittende eieren los te maken. Breng de eieren met zo min mogelijk meel over in een uitgedeeld klein zeefje, eventueel met behulp van een penseeltje. Omdat het aantal eieren laag is, is het belangrijk om alle eieren in het zeefje te krijgen. Maak nu twee petrischaaltjes: ééntje met een laagje kraanwater en ééntje met een laagje bleek.

Draag een labjas om je kleren te beschermen tegen bleek en handschoenen om jezelf te beschermen tegen het bijtende melkzuur!

Was de eieren eerst in het schaaltje met water. Plaats het zeefje met eieren 4 minuten in het petrischaaltje met bleek. Beweeg het zeefje af en toe een beetje of pipetteer wat bleek uit het petrischaaltje weer bovenop de eieren. Daarna meteen weer in het schaaltje met kraanwater plaatsen en het bleek goed wegwassen. Probeer de eieren een beetje op een hoopje te krijgen. Droog het zeefje wat met papier en breng de eieren of larfjes met een penseeltje over op een objectglaasje. Druppel er een klein beetje melkzuur bovenop en dek het geheel af met een dekglaasje. Als je veel eieren hebt: niet teveel eieren op het objectglaasje doen; liever een paar meer preparaten maken! Omdat er waarschijnlijk weinig eieren zijn, is het belangrijk om alle eieren op objectglaasjes te krijgen. Gebruik de stereomicroscoop. De preparaten komen overnacht op 60°C, zodat het melkzuur goed kan inwerken.

Gebruikte petrischaaltjes mogen in de prullenbak. Zeefjes s.v.p. afgespoeld weer inleveren

Preparaten bekijken

Je kunt nu de preparaten onder de lichtmicroscoop bekijken. Maak foto's. Kun je erachter komen welke groep tegen welk gen / welke genen RNAi heeft gedaan? Lees Hoofdstuk 11 van het boek Evolution van Barton et al (Brightspace Week 6, 10 Oktober)

Elke student levert een eigen practicumverslag in op Brightspace.

Practicum: Fossils

Answer the 14 questions below, you can add pictures of your drawings. You can form teams of two and work together. Remember that answers to the questions need to be compiled in a single PDF document and **uploaded in Brightspace by each student individually**. Make sure you clearly state the number on the box of your fossils on your report. First take all the fossils out of the box and place them on the table.

Taphonomy

Sort the fossils according to their preservation type we had a look at in the lectures and discuss within your team.

1. Write down which preservation modes are represented. Write down examples you find in the box of fossils.

2. What is preserved of the organism and what could you reconstruct from the fossil how the organism looked like and how it lived? Give an answer for the following fossils:

- Plant fossil(s)
- Vertebrate
- Belemnite
- Ammonite
- Bivalve
- Sponge
- Brachiopod
- Coral

3. How much of the original organism is preserved? Make an estimation for the following fossil specimens:

- Plant fossil(s)
- Vertebrate
- Belemnite
- Ammonite
- Bivalve
- Sponge
- Brachiopod
- Coral

Phylogeny, taxonomy, comparative anatomy

4. Sort your fossils according to their phylogenetic position. Which classes are present?

5. Choose all the corals and draw them with the emphasis on the important characters and label them.

6. Choose all the cephalopods and draw them with the emphasis on the important characters and label them.

7. Draw the trilobite and label the important characters.

8. Draw a bivalve and a brachiopod from the box of fossils, label them and explain the differences.

Habitat and ecology

9. Sort your fossils according to the habitat they probably lived in. Write down the habitats you think are represented, write down examples you find in the box of fossils.

10. Have a look at the bivalves, brachiopods and echinoderms, which lifestyle did they have? In the sediment (infaunal), on the sediment (epifaunal), sessile, free-swimming, etc.? Write down the lifestyle and write down the example you find in the box of fossils.

Stratigraphy

11. Sort your fossils according to the three Phanerozoic erathem / era that are represented in the box of fossils. Can you name the erathem / era we discussed in the lectures?

12. Which corals and arthropods you have fossils of got extinct at the Permian - Triassic mass extinction?

13. Which vertebrates and cephalopods you have fossils of got extinct at the Cretaceous boundary?

14. Which fossil is the coolest? Make a drawing and explain why you think this is the coolest fossil.