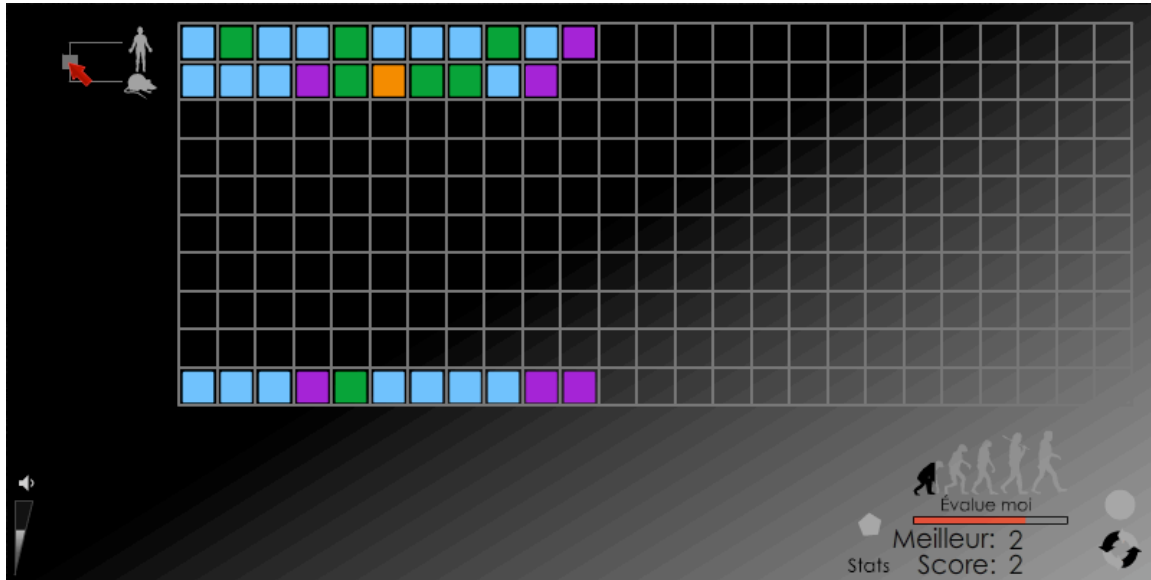


Tutorial

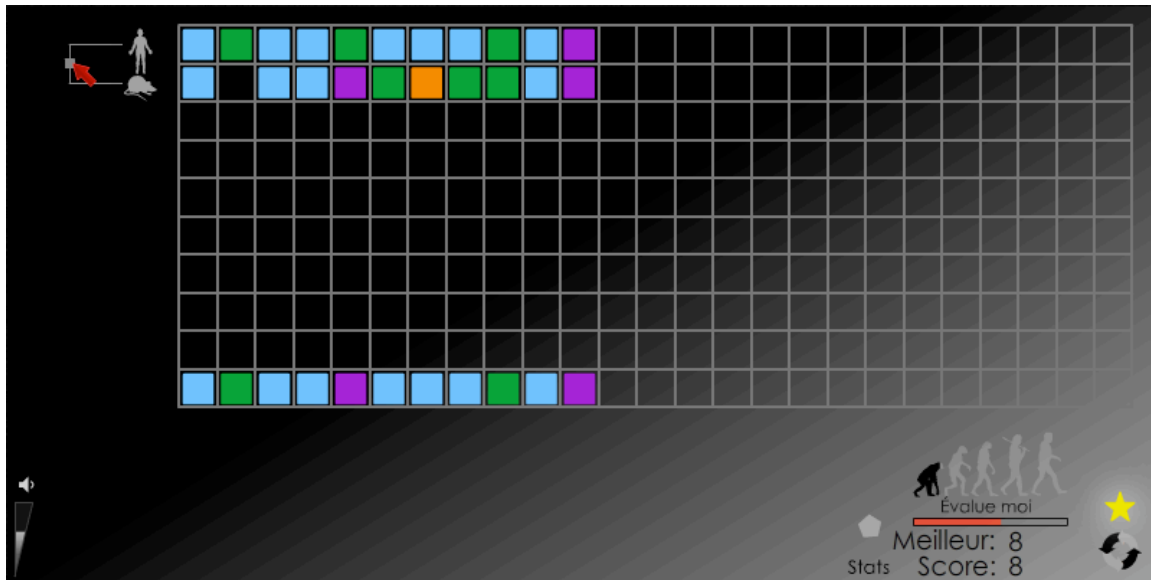
This tutorial shows how are computed the scores obtained by Phylo, and gives some tricks to complete a puzzle. The level id of this puzzle is 1946.

Stage 1:

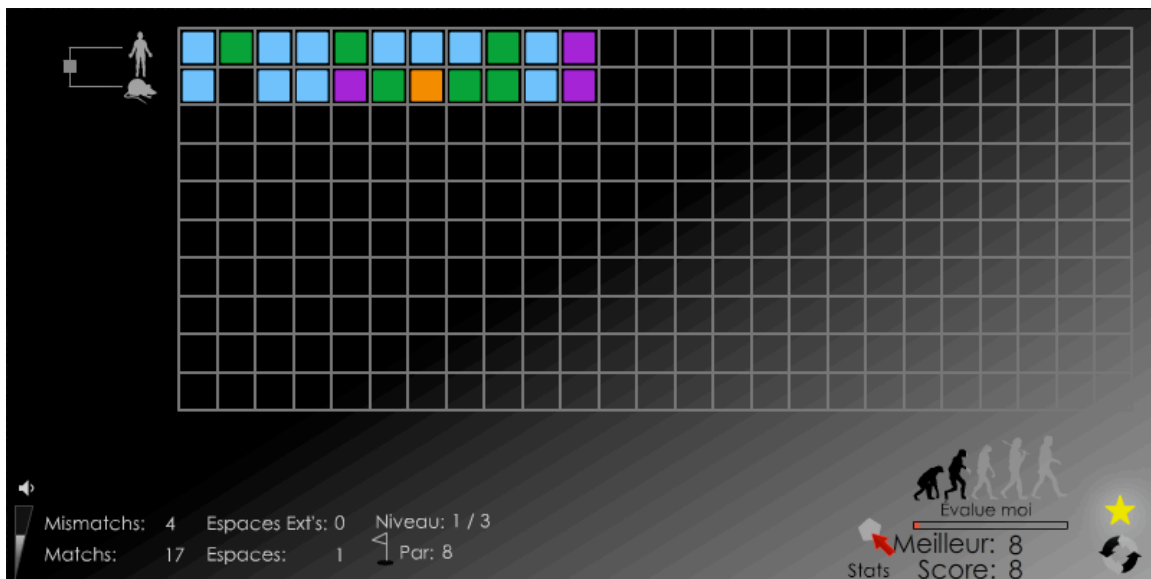


Push everything on the left and click on the father node to look at the ancestral sequence. When you compare the ancestor to the sequence on the top you have 7 matches and 4 mismatches. Now, if you compare the ancestor to the bottom sequence you have 7 matches 3 mismatches and 1 indel. A match brings a bonus of +1, while a mismatch costs -1 and an indel has a penalty equal to 4 plus the size of the gap. Thus, since this conformation has a total of 14 matches, 7 mismatches and 1 indel, we obtain a score of $14 \times (+1) + 7 \times (-1) + (-4 + 1 \times (-1)) = 2$.

Is it the best we can do? Apparently no. The star is not illuminated. How can you improve that?

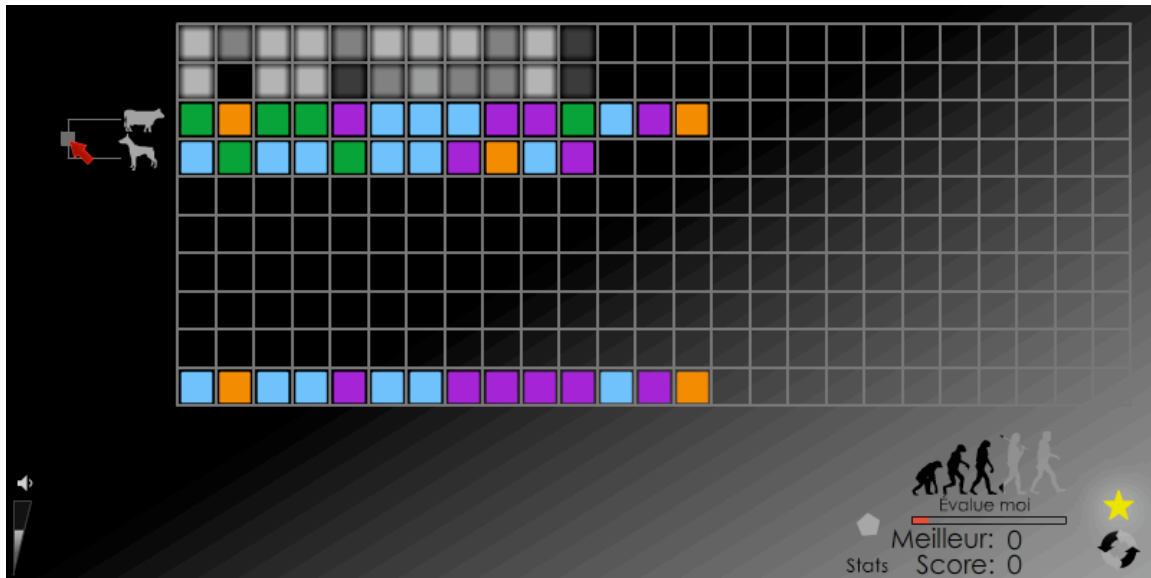


Shift now all nucleotides except the first one of one square to the right. You created a gap at the second position but significantly improved the number of matches between with the ancestral sequence. You have now 10 matches and 1 mismatch with the top sequence, and 7 matches, 3 mismatches and 1 indel for the bottom sequence. This makes a score of $17 - 4 - 5 = 8$! You can retrieve this detailed count if you click on the pentagon:

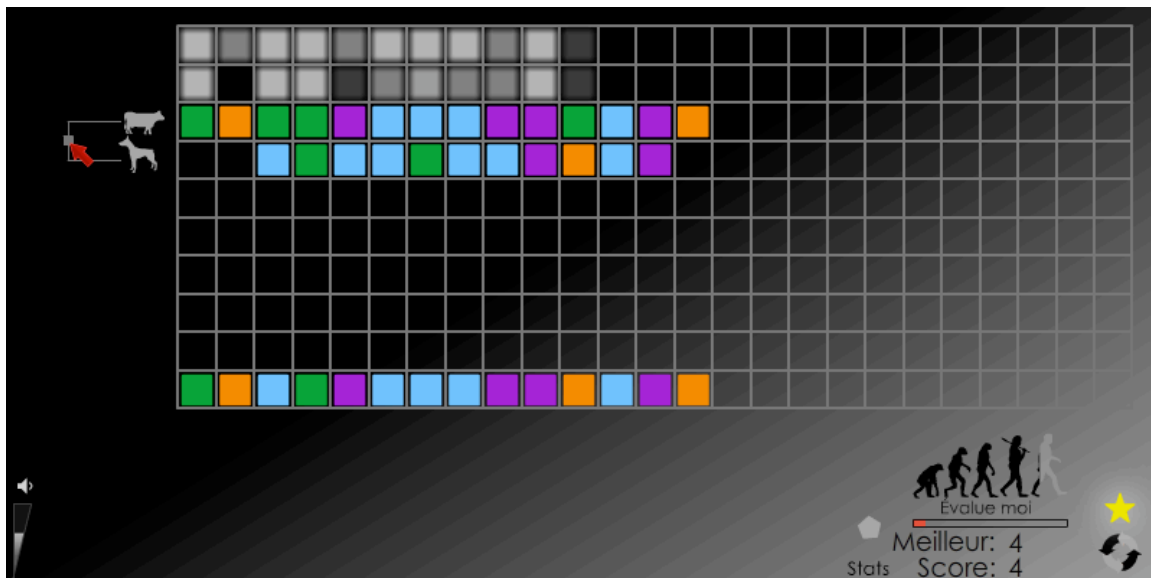


The star is now illuminated. Click on it and go to the next stage.

Stage 2:



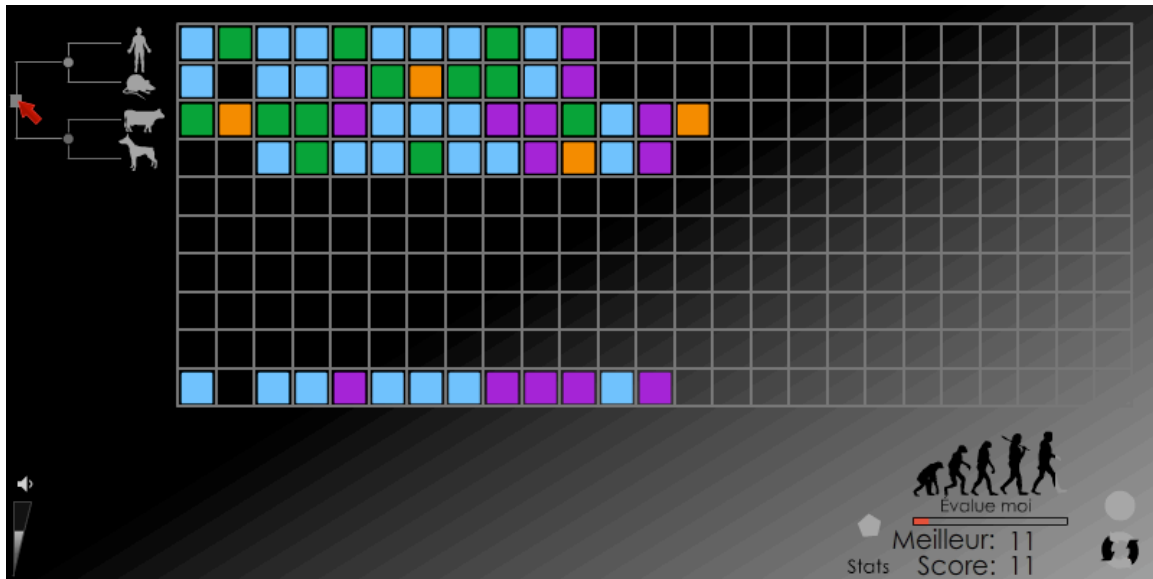
On more time, lets start by pushing everything to the left. You proceed exactly as above and your total count is here of 16 matches, 9 mismatches and 1 indel of size 3. Your score is thus $16 - 9 - 7 = 0$. The star is illuminated and you can go to the next stage. Can you do better though? This configuration does not look optimal...



Shift the bottom sequence of 2 positions to the left. Your score is now 4! Why? The ancestral sequence has changed and you have 20 matches (12 on top and 8 at the bottom), 5 mismatches (2 on top and 3 at the bottom) and 1 indel of size 2 and another indel of size 1 (bottom sequence). Thus, Your detailed count is $20 - 5 - (6 + 5) = 4$.

You improved the par (the computer's score was 0) and since your score seems optimal, just go to the next stage now. Click on the star!

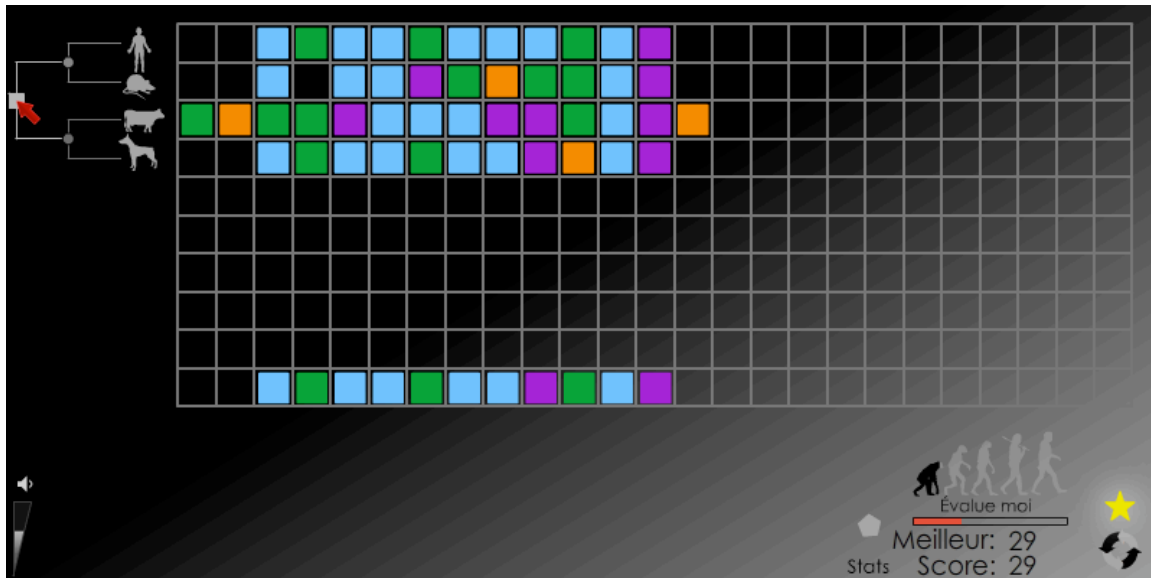
Stage 3:



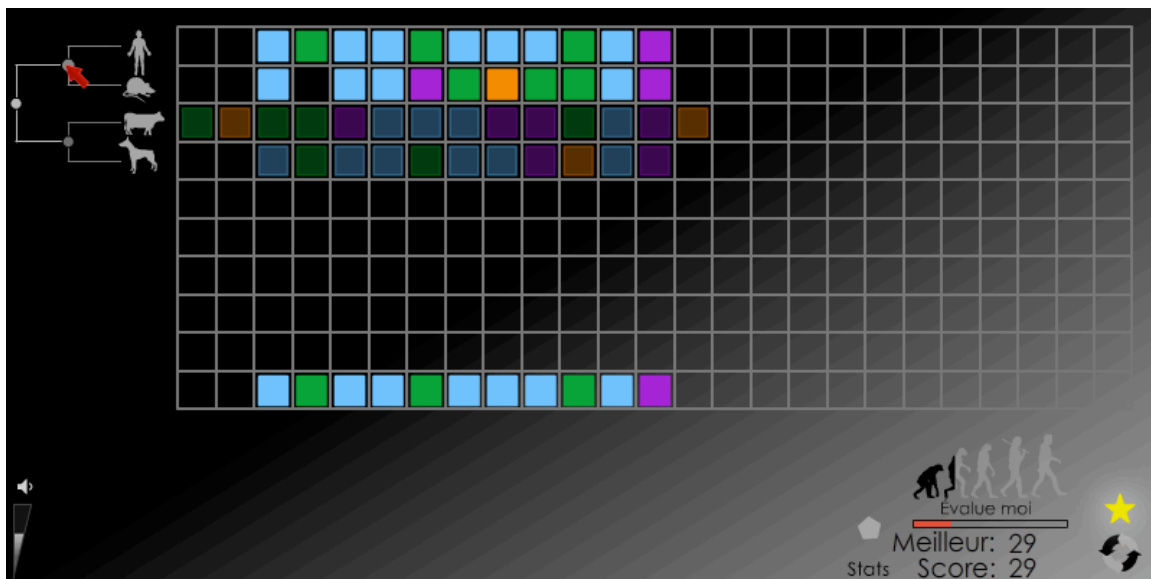
You must now assemble the best solutions previously found. If you just stack the best configuration of the two first sequences with the best configuration for the two last sequences without changing their position, then you obtain a score of 11, which is not enough to beat the par. Look at the ancestor to have an idea how far you are.

This configuration does not look coherent. How can we improve it?

Intuitively, we want to align, as much as possible, the 1st block (the two top sequences) and the 2nd block (the two lower sequences), and create the most compact configuration. You can also note that the color of the two last columns of the 1st block match the color of the 2nd and 3rd columns before the last of the 2nd block. You may want to align them too.



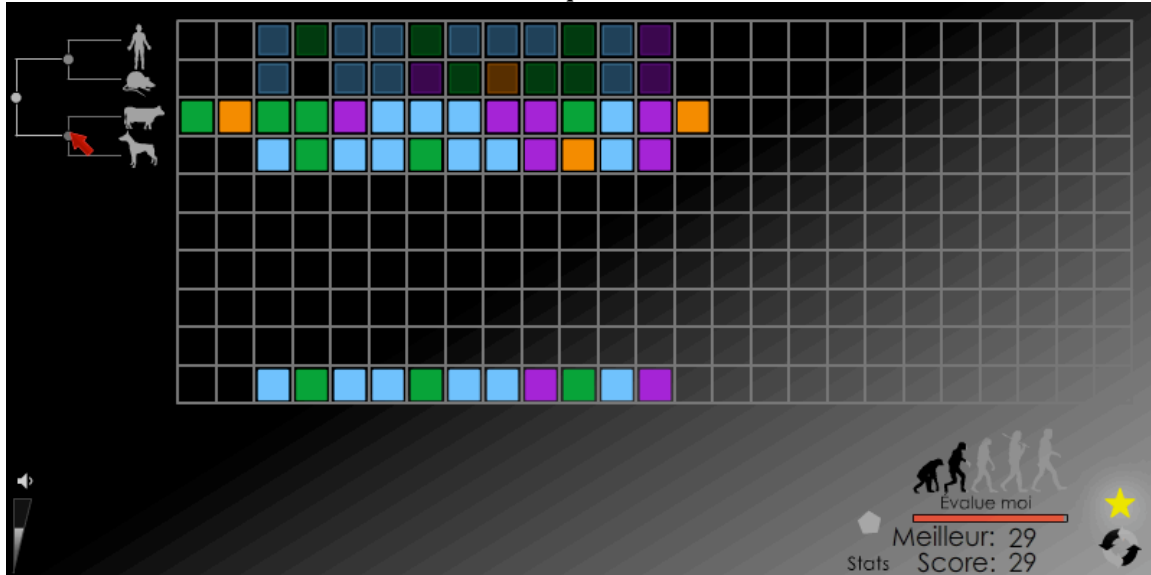
If you shift the 1st block by 2 positions to the left, you build a nice central block, which also reduce the number of different gaps. In addition you also created more full single color columns with the maximal height. How do you score this configuration?



Look at the intermediate ancestors once again and compare them with their child and father nodes. Click at the ancestor of the two first sequences. It is mostly identical, but shifted of two positions. Only the nucleotide at the 5th position changed but since this position is ambiguous (green or purple produce the same score), this ancestor is equivalent to the one used at the 1st stage. The score of this ancestor with the two sequences remains the same and is equal to 8.

However, we need now to also score the alignment between this ancestral sequence and the one at the root of the tree. Here, you have 10 matches and 1 mismatch and a score of 9. The total score from the root to the two first sequences is $8 + 9 = 17$.

Now, let's have a look to the two lower sequences.



Click on the immediate ancestor to the two last sequences. The ancestral sequence drastically changed! If you compute the score of this sequence with its two children, you obtain a score of 1 because you have now 17 matches (7 on the top and 10 at the bottom), 5 mismatches (4 on the top and 1 at the bottom), and 2 indels of size 2 and 1 ($17 - 5 - 6 - 5 = 1$). Which is lower than the one obtained before. Why that?

If you compare this intermediate ancestor with the ancestor at the root of the tree, you can see that both sequences match perfectly and thus you have 11 matches and a score of 11 on this edge. In fact, the ancestor sequence has been updated to optimize the total score along all paths from the root to the leaves.

It follows that the total score of this configuration is 8 (root to ancestor of the two first sequences) + 11 (root to ancestor of the two last sequences) + 9 (the two first sequences and their ancestor) + 1 (the two last sequences and their ancestor) = 29!

You have the highest score and you can submit it!