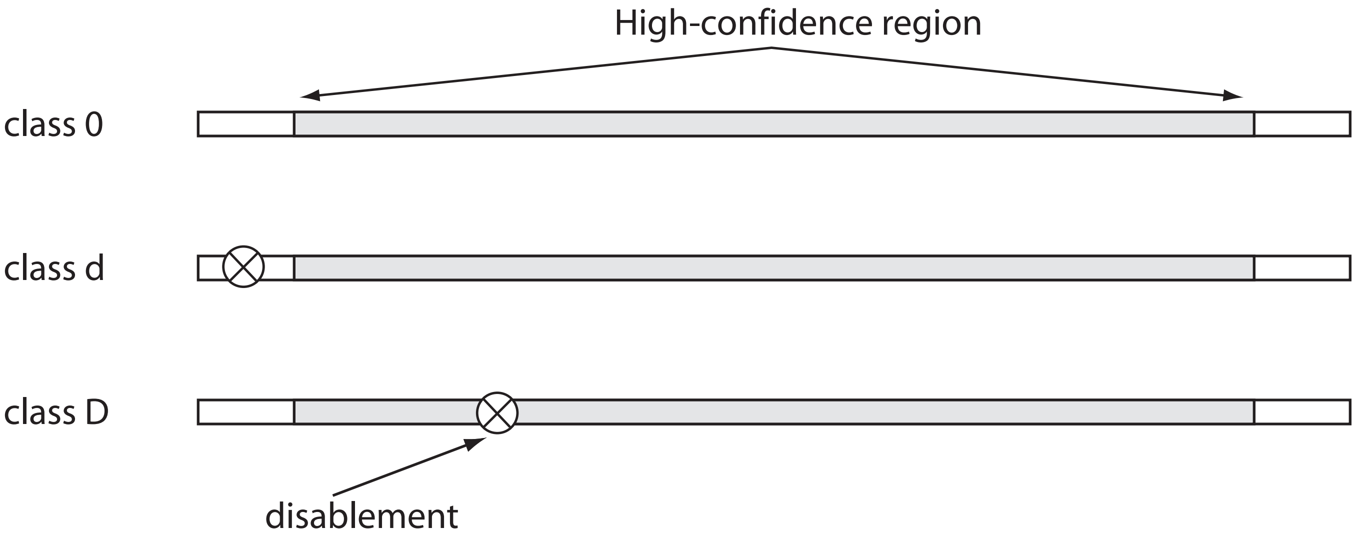
This graph illustrates the procedures described in the MATERIALS AND METHODS (page 30) on checking for frame disruptions in the pseudogene sequences.

…

Checking for Frame Disruptions

Sequence alignment can introduce artificial disablement at the ends of the sequence as mentioned above. We checked the existence of disablement for each pseudogene candidate, and put them into three classes: “0”, “d” or “D” according to whether the sequence contains disablement and where the disablements occur. For any pseudogene candidate, we designate a “high-confidence region” by scanning from both ends of the sequence until we have reached a position on the sequence where the amino acid sequence identity is greater than 50% for the next 10 residues and greater than 66.7% for the next three residues. We label a pseudogene candidate as “0” if no disablement is detected on the entire sequence, and label the sequence as “d” if disablement is detected outside the “high-confidence” region. A sequence is labeled “D’ if disablement is detected in the “high-confidence” region. The sequences that were labeled “d” were visually examined to determine whether the disablements are artifacts. The reason for such meticulous treatment is because we want to avoid introducing artificial disablement from the sequence alignment output. Note it is possible that some processed pseudogenes that were recently inserted into genome may not contain obvious disablements in their coding region; nevertheless they are disabled because of lack of promoter sequence.

…


High-confidence region
class 0
class d
class D
disablement