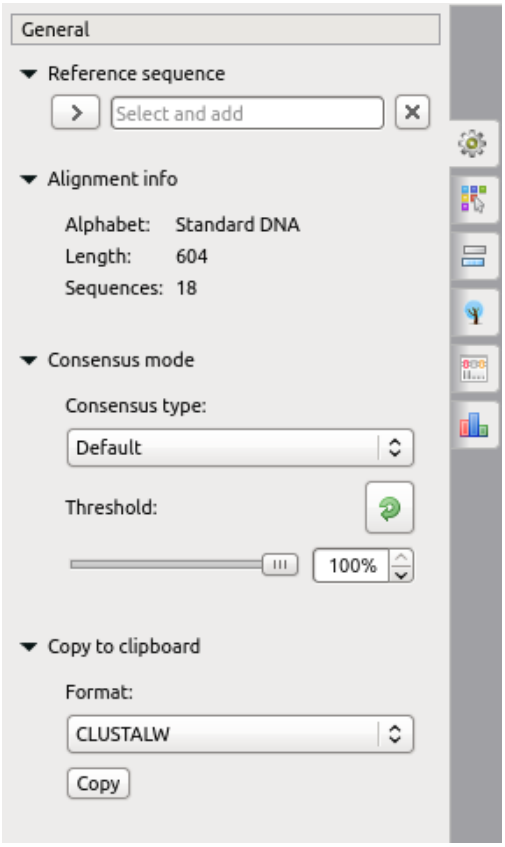
Consensus

Each base of a consensus sequence is calculated as a function of the corresponding column bases (and, for some algorithms, the column bases of the whole alignment). There are different methods to calculate the consensus. Each method reveals unique biological properties of the aligned sequences. The *Alignment Editor* allows switching between different consensus modes. To switch the consensus mode go to the *General tab* of the *Options Panel* or activate the context menu (using the right mouse button) or the *Actions* menu and select the *Consensus mode* item and *General tab* will be opened automatically:



There are several consensus modes:

- **Default** — the mode is based on the JalView algorithm. A character in the consensus is calculated based on the characters in the corresponding column and the value of the threshold:
  - If the percentage value of a most frequent character is greater than the threshold, the character is shown in upper case in the consensus.
  - If there are two characters with the equal highest frequency in a column, the '+" sign is shown.
  - If the percentage value of a most frequent character is high, but lower than the specified threshold, the character is shown in lower case in the consensus.
- **ClustalW** — emulates the ClustalW program behavior.
  - If the percentage value of a most frequent character is equal to 100%, the ***** base is shown in the consensus.
  - Otherwise, no characters are shown.
- **Levitsky** (not for AMINO alignment) — proposed by Victor Levitsky, this mode calculates the consensus of a DNA alignment, taking into account frequencies of characters in the whole alignment, i.e. not only one column. The algorithm includes these steps:
  1. Collect global alignment frequencies for every character in alignment using the extended DNA alphabet. For example:
     The **A** base increments the counter of the following bases:
     - A (the base itself),
     - W (A or T),
     - R (A or G),
     - M (A or C),
     - V (A or C or G),
     - H (A or C or T),
     - D (A or G or T),
     - N (A or C or G or T).
  2. Check the frequency of each symbol in the column:
     - If the frequency percentage value is is greater than the threshold, this base is considerer as a cosequence symbol - see further algorithm description.
     - If the frequency percentage value is lower than the specified threshold, this base is **not** considered as a cosequence symbol.
  3. from the set of bases, not filered on the point **2**, choose the rarest characters in the whole alignment (global frequencies have been calulated on the point **1**):
     - If there is only one base, this character is shown in the consensus.
     - If there are two or more bases, these symbols should be joint into one common base. For example, if there are two bases **A** and **T**, they will be joint into an extened DNA alphabet base **W**. Check this table to find out all possible combinations.

- **Simple extended** (not for AMINO alignment) — a character of the consensus is calculated based on the characters in the corresponding column. This algorithm joints all bases in the column (but only the ones which frequency percentage value is greater than the threshold) into a single base from the DNA extended alphabet using using this table.
- **Strict** — if a character percentage in a column is higher than the specified threshold, the character is shown in the consensus at this position. Otherwise, a gap character ('—') is shown.

Also the *General tab* shows the general information about an alignment and allows one to set up a reference sequence. The following chapter describes how to export a consensus sequence:

- Export Consensus