

Claims

1. A method for identifying a multi-copy region in a microorganism target fragment, comprising at least the following operations:

S100, searching for a candidate multi-copy region: performing an internal alignment on a microorganism target fragment, and searching for a region corresponding to a to-be-detected sequence of which a similarity meets a preset value as the candidate multi-copy region, the similarity being a product of a coverage rate and a matching rate of the to-be-detected sequence; wherein in the internal alignment, each two to-be-aligned sequences are separated in the microorganism target fragment, and positions of bases between each two to-be-aligned sequences do not cross;

S200, verifying the candidate multi-copy region to obtain the multi-copy region: obtaining a median value of copy numbers of the candidate multi-copy region; if the median value of the copy numbers of the candidate multi-copy region is greater than 1, the candidate multi-copy region is recorded as a multi-copy region; wherein the median value of copy numbers is a value positioned in the middle of a variable series formed by arranging copy numbers of each base position within the candidate multi-copy region in the statistical population in order of value size.

2. The method for identifying a multi-copy region in a microorganism target fragment according to claim 1, further comprising one or more of the followings:

- a. the coverage rate = (length of similar sequence / (end value of the to-be-detected sequence - starting value of the to-be-detected sequence + 1))%;
- b. the preset value of similarity exceeds 80%;
- c. the method further comprises the following operations: S101, aligning selected adjacent microorganism target fragments in pairs; if the similarity after alignment is lower than the preset value, issuing an alarm and displaying screening conditions corresponding to a target strain;
- d. in operation S200, a 95% confidence interval of the copy numbers of the candidate multi-copy region is calculated.

3. The method for identifying a multi-copy region in a microorganism target fragment according to claim 2, wherein when calculating the 95% confidence interval of the copy numbers of the candidate multi-copy region, a base number of the candidate multi-copy region serves as a sample number, and a copy number value corresponding to each base in the candidate multi-copy region serves as a sample value for calculation.

4. The method for identifying a multi-copy region in a microorganism target fragment according to claim 1, wherein when the microorganism target fragment includes multiple incomplete motifs, the motifs are connected together before searching for the candidate multi copy region.

5. The method for identifying a multi-copy region in a microorganism target fragment according to claim 3, further comprising one or more of the followings:

- a. if a region where the similarity meets the preset value contains different motifs, the region is cut based on an original motif connection point and divided into two regions, to determine whether the two regions are candidate multi-copy regions, respectively;
- b. the motifs are connected in any order.

6. The method for identifying a multi-copy region in a microorganism target fragment according to claim 1, wherein the obtaining of the median value of the copy numbers of the candidate multi-copy region includes: determining a position of each candidate multi-copy region on the microorganism target fragment, obtaining a number of other candidate multi-copy regions covering a position of each base of the to-be-verified candidate multi-copy region, and calculating the median value of the copy numbers of the to-be-verified candidate multi-copy region.

7. A device for identifying a multi-copy region in a microorganism target fragment, comprising at least the followings:

a candidate multi-copy region searching module, configured to perform internal alignment on a microorganism target fragment, and search for a region corresponding to a to-be-detected sequence of which a similarity meets a preset value as a candidate multi-copy region, the similarity being a product of a coverage rate and a matching rate of the to-be-detected sequence; wherein in the internal alignment, each two to-be-aligned sequences are separated in the microorganism target fragment, and positions of bases between each two to-be-aligned sequences do not cross;

a candidate multi-copy region verifying module, configured to obtain a median value of copy numbers of the candidate multi-copy region; if the median value of the copy numbers of the candidate multi-copy region is greater than 1, the candidate multi-copy region is recorded as a multi-copy region.

8. The device for identifying a multi-copy region in a microorganism target fragment according to claim 7, further comprising one or more of the followings:

- a. the coverage rate = (length of similar sequence/ (end value of the to-be-detected sequence - starting value of the to-be-detected sequence + 1))%;

- b. the preset value of similarity exceeds 80%;
- c. the candidate multi-copy region searching module further includes the following submodules:
 - a raw data similarity comparison module, configured to align selected adjacent microorganism target fragments in pairs; if the similarity after alignment is lower than the preset value, an alarm is issued and screening conditions corresponding to a target strain are displayed;
- d. the candidate multi-copy region verifying module is further configured to calculate a 95% confidence interval of the copy numbers of the candidate multi-copy region.

9. The device for identifying a multi-copy region in a microorganism target fragment according to claim 8, wherein when calculating the 95% confidence interval of the copy numbers of the candidate multi-copy region, a base number of the candidate multi-copy region serves as a sample number, and a copy number value corresponding to each base in the candidate multi-copy region serves as a sample value for calculation.

10. The device for identifying a multi-copy region in a microorganism target fragment according to claim 7, wherein in the candidate multi-copy region searching module, when the microorganism target fragment includes multiple incomplete motifs, the motifs are connected together before searching for the candidate multi-copy region.

11. The device for identifying a multi-copy region in a microorganism target fragment according to claim 10, further comprising one or more of the followings:

- a. if a region where the similarity meets the preset value contains different motifs, the region is cut based on an original motif connection point and divided into two regions, to determine whether the two regions are candidate multi-copy regions, respectively;
- b. the motifs are connected in any order.

12. The device for identifying a multi-copy region in a microorganism target fragment according to claim 7, wherein the candidate multi-copy region verifying module further includes a candidate multi-copy region copy number median value obtaining submodule, to determine a position of each candidate multi-copy region on the microorganism target fragment, obtain a number of other candidate multi-copy regions covering a position of each base of the to-be-verified candidate multi-copy region, and calculate the median value of the copy numbers of the to-be-verified candidate multi-copy region.

13. A computer readable storage medium, which stores a computer program, wherein when executed by a processor, the program implements the method for identifying a multi-copy region in a microorganism target fragment according to any one of claims 1-6.

14. A computer processing device, comprising a processor and the computer readable storage medium according to claim 13, wherein the processor executes the computer program on the computer readable storage medium to implement the operations of the method for identifying a multi-copy region in a microorganism target fragment according to any one of claims 1-6.

15. An electronic terminal, comprising a processor, a memory and a communicator; wherein the memory stores a computer program, the communicator communicates with an external device, and the processor executes the computer program stored in the memory, so that the electronic terminal executes the method for identifying a multi-copy region in a microorganism target fragment according to any one of claims 1-6.

16. A use of the method for identifying a multi-copy region in a microorganism target fragment according to any one of claims 1-6, the device for identifying a multi-copy region in a microorganism target fragment according to any one of claims 7-12, the computer readable storage medium according to claim 13, the computer processing device according to claim 14, or the electronic terminal according to claim 15 in the detection of a multi-copy region in a microorganism target fragment.

17. The use according to claim 16, wherein the microorganism includes one or more of bacterium, virus, fungus, amoeba, cryptosporidium, flagellate, microsporidium, piroplasma, plasmodium, toxoplasma, trichomonas, and kinetoplastid.