

Claims

1. A method for identifying a specific region in a microorganism target fragment, at least comprising:

performing PCR amplification of a specific region in a target fragment of the microorganism, wherein a process for identifying the specific region at least comprises:

step S100, respectively comparing a microorganism target fragment with whole genome sequences of a first plurality of comparison strains one-to-one, and then cutting and removing fragments in the microorganism target fragment of which a similarity exceeds a preset value, to obtain a plurality of residual fragments as first-round cut fragments T_1 - T_n , wherein n is an integer greater than or equal to 1;

step S200, respectively comparing the first-round cut fragments T_1 - T_n with whole genome sequences of a second plurality of comparison strains by group iteration, and then cutting and removing fragments in the first-round cut fragments T_1 - T_n of which the similarity exceeds a preset value, to obtain a collection of residual cut fragments as a candidate specific region of the microorganism target fragment; and

step S300, verifying the candidate specific region: determining whether the candidate specific region meets the following requirements:

1) searching in public databases to find whether there are other species of which a similarity to the candidate specific region is greater than the preset value;

2) respectively comparing the candidate specific region with whole genome sequences of the first plurality of and the second plurality of comparison strains and a whole genome sequence of a host of a source strain of the microorganism target fragment, to find whether there are fragments with a similarity greater than the preset value;

if the candidate specific region does not meet the above requirements, the candidate specific region is a specific region of the microorganism target fragment.

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

a. the similarity refers to a product of a coverage rate and a matching rate of the microorganism target fragment, and the coverage rate = (length of similar sequence fragment / (end value of the microorganism target fragment - starting value of the

microorganism target fragment + 1))%;

- b. the method further comprises the following operation: step S110, comparing selected adjacent microorganism target fragments in pairs; if the similarity after comparison is lower than the preset value, issuing an alarm and displaying screening conditions corresponding to a target strain.

3. The method for identifying a specific region in a microorganism target fragment according to claim 1, wherein the first-round cut fragment T_n being compared with whole genome sequences of the second plurality of comparison strains by group iteration includes the following operations:

step S210, dividing the second plurality of comparison strains into P groups, each group including a third plurality of comparison strains;

step S211, simultaneously comparing the first-round cut fragment T_n with the whole genome sequences of each comparison strain in the first group one-to-one, and removing fragments of which the similarity exceeds the preset value, to obtain a plurality of residual fragments as a first-round candidate sequence library of the first-round cut fragment T_n ; and

step S212, simultaneously comparing a previous-round candidate sequence library of the first-round cut fragment T_n with whole genome sequences of each comparison strain in the next group one-to-one, and removing fragments of which the similarity exceeds the preset value, to obtain a plurality of residual fragments as a next-round candidate sequence library of the first-round cut fragment T_n ; repeating operation step S212 from the first-round candidate sequence library until a Pth-round candidate sequence library is obtained as a candidate specific sequence library of the first-round cut fragment T_n ;

wherein a collection of all the candidate specific sequence libraries of the first-round cut fragments is the candidate specific region.

4. The method for identifying a specific region in a microorganism target fragment according to claim 3, wherein a number of comparison strains contained in a comparison strain group is set according to a hardware configuration of a computing environment.

5. The method for identifying a specific region in a microorganism target fragment according to

claim 1, further comprising: after operation step S200,

performing operations step S100 and step S200 to obtain candidate specific regions of each microorganism target fragment in a target sequence, taking a collection of the candidate specific regions of each microorganism target fragment as candidate specific regions of the target sequence.

6. A device for identifying a specific region in a microorganism target fragment, comprising:

a first-round cut fragment obtaining module, configured to respectively compare a microorganism target fragment with whole genome sequences of a first plurality of comparison strains one-to-one, and then cut and remove fragments of which a similarity exceeds a preset value, to obtain a plurality of residual fragments as first-round cut fragments T_1 - T_n , wherein n is an integer greater than or equal to 1;

a candidate specific region obtaining module, configured to respectively compare the first-round cut fragments T_1 - T_n with whole genome sequences of a second plurality of comparison strains by group iteration, and then cut and remove fragments in the first-round cut fragments T_1 - T_n of which the similarity exceeds a preset value, to obtain a collection of residual cut fragments as candidate specific regions of the microorganism target fragment; and

a candidate specific region verifying module, configured to determine whether the candidate specific region meets the following requirements:

1) searching in public databases to find whether there are other species of which a similarity to the candidate specific region is greater than the preset value;

2) respectively comparing the candidate specific region with whole genome sequences of the first plurality of and the second plurality of comparison strains and a whole genome sequence of a host of a source strain of the microorganism target fragment, to find whether there are fragments with a similarity greater than the preset value;

if the candidate specific region does not meet the above requirements, the candidate specific region is a specific region of the microorganism target fragment.

7. The device for identifying a specific region in a microorganism target fragment according to claim 6, further comprising one or more of the following:

a. the similarity refers to a product of a coverage rate and a matching rate of the

microorganism target fragment, and the coverage rate = (length of similar sequence fragment/(end value of the microorganism target fragment - starting value of the microorganism target fragment + 1))%;

- b. the first-round cut fragment obtaining module further includes the following submodules: a raw data similarity comparison submodule, configured to compare selected adjacent microorganism target fragments in pairs; if the similarity after comparison is lower than the preset value, an alarm is issued and screening conditions corresponding to the target strain are displayed.

8. The device for identifying a specific region in a microorganism target fragment according to claim 6, wherein when the first-round cut fragment T_n is compared with whole genome sequences of the second plurality of comparison strains by group iteration, the candidate specific region obtaining module includes the following submodules:

a comparison strain grouping submodule, configured to divide the second plurality of comparison strains into P groups, each group including a third plurality of comparison strains;

a first-round candidate sequence library obtaining submodule, configured to simultaneously compare the first-round cut fragment T_n with the whole genome sequences of each comparison strain in the first group one-to-one, and remove fragments of which the similarity exceeds the preset value, to obtain a plurality of residual fragments as a first-round candidate sequence library of the first-round cut fragment T_n ;

a candidate specific region obtaining submodule, to simultaneously compare a previous-round candidate sequence library of the first-round cut fragment T_n with whole genome sequences of each comparison strain in a next group one-to-one, and remove fragments of which the similarity exceeds the preset value, to obtain a plurality of residual fragments as a next-round candidate sequence library of the first-round cut fragment T_n ; the candidate specific region obtaining submodule is repeated from the first-round candidate sequence library until a Pth-round candidate sequence library is obtained as a candidate specific sequence library of the first-round cut fragment T_n ;

wherein a collection of all the candidate specific sequence libraries of the first-round cut fragments is the candidate specific region.

9. The device for identifying a specific region in a microorganism target fragment according to claim 8, wherein a number of comparison strains contained in a comparison strain group is set according to a hardware configuration of a computing environment.

10. The device for identifying a specific region in a microorganism target fragment according to claim 6, further comprising:

a candidate specific region obtaining module of a target sequence, which executes the first-round cut fragment obtaining module of a target fragment and the candidate specific region obtaining module of a microorganism target fragment to obtain candidate specific regions of each microorganism target fragment in the target sequence, and takes a collection of the candidate specific regions of each microorganism target fragment as candidate specific regions of the target sequence.

11. A computer readable storage medium, which stores a computer program, wherein when executed by a processor, the computer program implements the method for identifying a specific region in a microorganism target fragment according to any one of claims 1-5.

12. A computer processing device, comprising a processor and the computer readable storage medium according to claim 11, wherein the processor executes the computer program on the computer readable storage medium to implement the method for identifying a specific region in a microorganism target fragment according to any one of claims 1-5.

13. An electronic terminal, comprising a processor, a memory and a communicator; the memory stores a computer program, the communicator communicates with an external device, and the processor executes a computer program stored in the memory, so that the electronic terminal executes the method for identifying a specific region in a microorganism target fragment according to any one of claims 1-5.

14. A use of the method for identifying a specific region in a microorganism target fragment according to any one of claims 1-5, the device for identifying a specific region in a microorganism target fragment according to any one of claims 6-10, the computer readable

storage medium according to claim 11, the computer processing device according to claim 12, or the electronic terminal according to claim 13 for identifying a specific region in a microorganism target fragment.

15. The use according to claim 14, further comprising one or more of the following:

- a. the microorganism includes one or more of bacterium, virus, fungus, amoeba, cryptosporidium, flagellate, microsporidium, piroplasma, plasmodium, toxoplasma, trichomonas and kinetoplastid.