CLAIMS

1. A method of defining microbial operational taxonomic units (OTUs) in a sample, the method comprising:

1) - obtaining a sample, which comprises microorganisms each of which comprises a phylogenetically information gene,

2) - obtaining raw sequence reads of the phylogenetically informative gene of the microorganisms in the sample using a PCR-based high-throughput sequencing technique,

3) - processing the raw sequence reads to obtain assembled, fully-length qualified sequences,

4) - obtaining a relative abundance value of each of the qualified sequences, wherein the total relative abundance of all qualified sequences is 100%;

5) - ranking from high to low all qualified sequences by their respective relative abundance value, and separating the qualified sequences into a high abundance group and a low abundance group, wherein the high abundance group consists of qualified sequences whose abundance values are higher than the those in the low abundance group and collectively account for about 70%-80% of the total abundance; and the low abundance group consists of the remaining qualified sequences which account for about 20%-30% of the total abundance;

6) - delineating OTUs in the sample using only qualified sequences in the high abundance group to obtain Tentative OTUs; and

7) - re-mapping qualified sequences in the low abundance group to the Tentative OTUs, and assigning them individually to a suitable Tentative OTUs only if they the qualified sequence has at least 90% sequence similarly to the OTU Sequence, to arrive at the final definition of OTUs.

2. A method of claim 1, wherein the phylogenetically informative gene is selected from the group consisting of the 16s rRNA gene or the 18s rRNA gene.

3. A method according to claim 2, wherein the phylogenetically informative gene is the 16s rRNA gene.

4. A method according to claim 3, wherein the phylogenetically informative gene is one or more variable regions of the 16s rRNA gene.

5. A method according to claim 4, wherein the one or more variable regions of the 16s rRNA gene is selected from the group consisting of the V3, V3-V4, V4, V5-V6, V9 hypervariable regions.

6. A method according to claim 1, wherein the raw sequence reads are obtained by filtering, quality-trimming, de-replicating and removing PCR primers to obtain qualified sequences.

7. The method of claim 1, wherein the OTU delineation is by a commercially available pipeline selected from the group consisting of USEARCH, QIIME, and MOTHUR.

8. The method of claim 1, wherein the DNA sequence is determined by a pyrosequencing method using an Illumina™ Sequencer.

9. The method of claim 1, wherein the total nucleic acid is isolated from the sample and then sequenced.

10. The method of claim 1, wherein Steps 4), 5), 6) and/or 7) are carried out by a processor.

11. The method of claim 1, wherein in Step 5) the high abundance group consists of qualified sequences whose abundance values are higher than the those in the low abundance group and collectively account for about 71%-79%, 72%-78%, 73%-77%, 74-76%, 74.5%-75.5%, 74.6%-75.4%, 74.7%-75.3%, 74.8%-75.2%, 74.9%-75.1% of the total abundance; and the low abundance group consists of the remaining qualified sequences which account for about 21%-29%, 22%-28%, 23%-27%, 24-26%, 24.5%-25.5%, 24.6%-25.5%, 24.7%-25.3%, 24.8%-25.2%, 24.9%-25.1% of the total abundance.

12. The method of claim 1, wherein in Step 5) the high abundance group consists of qualified sequences whose abundance values are higher than the those in the low abundance group and collectively account for about 75% of the total abundance; and the low abundance

group consists of the remaining qualified sequences which account for about 25% of the total abundance.

13. The method of claim 1, wherein in Step 7) assigning them individually to a suitable Tentative OTUs only if they the qualified sequence has at least 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, or 99% sequence similarly to the OTU Sequence.

14. A method for isolating a microorganism from an environmental sample, wherein the microorganism comprises a phylogenetically informative gene, the method comprising:

- determining OTUs in the environmental sample according to claim 1;
- selecting an OTU with its unique phylogenetically informative gene sequence as a to-be-isolated microorganism;
- culturing microorganisms in the sample;
- determining the DNA sequence of the phylogenetically informative gene of each of the cultured microorganisms; and
- isolating a microorganism the sequence of whose phylogenetically informative gene is homologous to the phylogenetically informative gene sequence of the to-be-isolated microorganism.

15. The method according to claim 14, wherein the microorganism is a bacterium.

16. The method according to claim 14, wherein a microorganism the sequence of whose phylogenetically informative gene is at least 95% identical to the phylogenetically informative gene sequence of the to-be-isolated microorganism is isolated.

17. The method according to claim 16, wherein a microorganism the sequence of whose phylogenetically informative gene is at least 99% identical to the phylogenetically informative gene sequence of the to-be-isolated microorganism is isolated.

18. The method according to claim 17, wherein a microorganism the sequence of whose phylogenetically informative gene is identical to the phylogenetically informative gene sequence of the to-be-isolated microorganism is isolated.

19. A method of defining microbial operational taxonomic units (OTUs) in a sample, the method comprising:

1) - obtaining a relative abundance value of each of qualified sequences of a phylogenetically informative gene in microorganisms contained in the sample, wherein the total relative abundance of all qualified sequences is 100%;

2) - ranking from high to low all qualified sequences by their respective relative abundance value, and separating the qualified sequences into a high abundance group and a low abundance group, wherein the high abundance group consists of qualified sequences whose abundance values are higher than the those in the low abundance group and collectively account for about 70%-80% of the total abundance; and the low abundance group consists of the remaining qualified sequences which account for about 20%-30% of the total abundance;

3) - delineating OTUs in the sample using only qualified sequences in the high abundance group to obtain Tentative OTUs; and

4) - re-mapping qualified sequences in the low abundance group to the Tentative OTUs, and assigning them individually to a suitable Tentative OTUs only if they the qualified sequence has at least 90% sequence similarly to the OTU Sequence, to arrive at the final definition of OTUs.

20. The method according to claim 19, wherein the qualified sequences are obtained by obtaining raw sequence reads of the phylogenetically informative gene of the microorganisms in the sample using a PCR-based high-throughput sequencing technique, and processing the raw sequence reads to obtain assembled, fully-length qualified sequences.

21. A method for identifying, characterizing or assessing a microbial community or microbiota in a sample, the method comprising:

1) - obtaining a relative abundance value of each of qualified sequences of a phylogenetically informative gene in microorganisms contained in the sample, wherein the total relative abundance of all qualified sequences is 100%;

2) - ranking from high to low all qualified sequences by their respective relative abundance value, and separating the qualified sequences into a high abundance group and a low abundance group, wherein the high abundance group consists of qualified sequences whose abundance values are higher than the those in the low abundance group and

collectively account for about 70%-80% of the total abundance; and the low abundance group consists of the remaining qualified sequences which account for about 20%-30% of the total abundance;

3) - delineating microbial operational taxonomic units (OTUs) in the sample using only qualified sequences in the high abundance group to obtain Tentative OTUs; and

4) - re-mapping qualified sequences in the low abundance group to the Tentative OTUs, and assigning them individually to a suitable Tentative OTUs only if they the qualified sequence has at least 90% sequence similarly to the OTU Sequence, to arrive at the final definition of OTUs.

22. A software executable by at least one processor to perform a method for defining microbial operational taxonomic units (OTUs) in a sample, the method comprising:

obtaining a relative abundance value of each of qualified sequences of a phylogenetically informative gene in microorganisms contained in the sample, wherein the total relative abundance of all qualified sequences is 100%;

ranking from high to low all qualified sequences by their respective relative abundance value, and separating the qualified sequences into a high abundance group and a low abundance group, wherein the high abundance group consists of qualified sequences whose abundance values are higher than the those in the low abundance group and collectively account for about 70%-80% of the total abundance; and the low abundance group consists of the remaining qualified sequences which account for about 20%-30% of the total abundance;

delineating OTUs in the sample using only qualified sequences in the high abundance group to obtain Tentative OTUs; and

re-mapping qualified sequences in the low abundance group to the Tentative OTUs, and assigning them individually to a suitable Tentative OTUs only if they the qualified sequence has at least 90% sequence similarly to the OTU Sequence, to arrive at the final definition of OTUs.

23. A system for defining microbial operational taxonomic units (OTUs) in a sample, the system comprising at least one processor configured to:

obtain a relative abundance value of each of qualified sequences of a phylogenetically informative gene in microorganisms contained in the sample, wherein the total relative abundance of all qualified sequences is 100%;

rank from high to low all qualified sequences by their respective relative abundance value, and separating the qualified sequences into a high abundance group and a low abundance group, wherein the high abundance group consists of qualified sequences whose abundance values are higher than the those in the low abundance group and collectively account for about 70%-80% of the total abundance; and the low abundance group consists of the remaining qualified sequences which account for about 20%-30% of the total abundance;

delineate OTUs in the sample using only qualified sequences in the high abundance group to obtain Tentative OTUs; and

re-map qualified sequences in the low abundance group to the Tentative OTUs, and assigning them individually to a suitable Tentative OTUs only if they the qualified sequence has at least 90% sequence similarly to the OTU Sequence, to arrive at the final definition of OTUs.