

CLAIMS

What is claimed is:

1. A method of detecting a pathogen in a biological sample, comprising:
receiving sequence data from a biological sample;
identifying k-mers in the sequence data that have an exact match in a hash table that is initialized with a first set of k-mers comprising pathogen k-mers in a genome of the pathogen and with a second set of k-mers comprising control k-mers; and
providing a detection output for the biological sample based at least in part on a first count of exact matches of the k-mers in the sequence data with the first set and a second count of exact matches of the k-mers in the sequence data with the second set, wherein the detection output comprises a positive result for a pathogen detection when both the first count is above a first set threshold and the second count is above a second set threshold, and wherein the detection output comprises a negative result for the pathogen detection when the first count is below the first set threshold, when the second count is below the second set threshold, or both.
2. The method of claim 1, wherein the k-mers in the sequence data, the first set, and the second set are of a fixed size that is greater than 24 nucleotides.
3. The method of claim 1, wherein the first set of k-mers is a subset of all k-mers in the genome of the pathogen.
4. The method of claim 3, wherein the subset is based on sufficient nonsimilarity to a control genome from which the control k-mers are derived.
5. The method of claim 4, wherein the control genome is a human genome.
6. The method of claim 1, wherein the first set of k-mers comprises variants in the genome of the pathogen.

7. The method of claim 1, wherein the first set is larger than the second set.
8. The method of claim 1, wherein the first set comprises k-mers from a plurality of different pathogens, and wherein the detection output for the biological sample comprises the pathogen detection of the pathogen of the plurality of different pathogens.
9. The method of claim 1, comprising aligning the sequence data to the genome of the pathogen based on the positive result for the pathogen detection.
10. The method of claim 9, comprising identifying sequence variants of the pathogen in aligned sequence data.
11. The method of claim 1, comprising administering a treatment for the pathogen responsive to the positive result for the pathogen detection.
12. A method of detecting a pathogen in a biological sample, comprising:
 - generating sequence data from a sequencing library prepared from a biological sample;
 - identifying k-mers in the sequence data that have an exact match in a hash table that is initialized with a set of k-mers comprising pathogen k-mers in a pathogen genome of a pathogen;
 - determining coverages in the sequence data for individual target regions of the pathogen genome based on one or both of a count of the identified k-mers or a number of sequence reads in the sequence data comprising the identified k-mers that correspond to the respective individual target regions in the pathogen genome, wherein an individual target region is determined to be covered when the count of the identified k-mers or the number of sequence reads that correspond to the individual target region is above a threshold count;

determining that a number of covered individual target regions is above a detection threshold; and

providing a detection output that the biological sample is positive for presence of the pathogen.

13. The method of claim 12, comprising identifying control k-mers in the sequence data that have an exact match with a control set of k-mers of a control genome.

14. The method of claim 13, comprising determining that a sufficient number of the individual target regions of the control genome have sufficient coverage based on the determined coverages.

15. The method of claim 13, comprising identifying sequence variants of the pathogen in the sequence data.

16. A sequencing device, comprising:
a substrate having loaded thereon a sequencing library prepared from a sample;
a computer programmed to:
 cause the sequencing device to generate sequence data from sequencing library;
 scan k-mers of a fixed size n in individual reads in the sequence data;
 access a hash table stored in a memory of the computer, the hash table being initialized with a set of reference k-mers of the fixed size n;
 identify exact matches of the k-mers with the set of reference k-mers using the hash table; and
 determine a characteristic of the sample based on a count of the identified exact matches being above a threshold.

17. The sequencing device of claim 16, comprising a field-programmable gate array that executes instructions of the programmed computer.
18. The sequencing device of claim 16, wherein the set of reference k-mers comprises k-mers of genomes of a plurality of gut microbes.
19. The sequencing device of claim 16, wherein the set of reference k-mers comprises k-mers of a pathogen panel comprising a plurality of pathogens.
20. The sequencing device of claim 16, wherein the set of reference k-mers comprises k-mers of a SARS-CoV-2 genome.