sequence construct. Disclosed herein is also a computer-implemented method of genotyping a genetic sample, comprising: providing a plurality of sequence reads corresponding to a genetic sample; aligning the plurality of sequence reads to a reference sequence construct that represents different genotypes as alternative sequences at a first position in the construct using a computer processor; and genotyping the sample based upon the alignment of the plurality of sequence reads against the reference sequence construct. Thus, in one aspect, the present invention refers to a method of genotyping a genetic sample, comprising obtaining a plurality of sequence reads corresponding to a genetic sample, wherein the plurality of sequence reads include portions of a structural variation; aligning the plurality of sequence reads to a reference directed acyclic graph (DAG) that accounts for variability in genetic sequences within a species, the reference DAG comprising a first path corresponding to a first allele and a second path corresponding to a second allele at a position in the reference DAG, wherein the first allele comprises a genetic structural variation, wherein the plurality of sequence reads that include portions of the genetic structural variation properly align to the reference DAG; and determining a genotype for the genetic sample based upon the number of sequence reads aligning to the first path and the second path, wherein the alignment of two or more related reads is used to determine the genotype, and the determined genotype comprises the structural variation. In another aspect, the present invention refers to a computer-implemented method of genotyping a genetic sample, comprising providing a plurality of sequence reads corresponding to a genetic sample, wherein the plurality of sequence reads include portions of a structural variation; aligning the plurality of sequence reads to a reference directed acyclic graph (DAG) that accounts for variability in genetic sequences within a species, the reference DAG comprising nodes, wherein the reference DAG comprises a first path corresponding to a first allele and a second path corresponding to a second allele at a position in the reference DAG using a computer processor, wherein the first allele comprises a genetic structural variation, wherein the plurality of sequence reads that include portions of the genetic structural variation properly align to the reference DAG; and determining a genotype for the genetic sample based upon the number of sequence reads aligning to the first path and the second path, wherein the alignment of two or more related reads is used to determine the genotype, and the determined genotype comprises the structural variation. In a further aspect, the present invention refers to a system for aligning a plurality of reads comprising a processor and memory, wherein the memory comprises instructions that, when executed, cause the processor to obtain a plurality of sequence reads corresponding to a genetic sample, wherein the plurality of sequence reads include portions of a structural variation; align the plurality of sequence reads to a reference directed acyclic graph (DAG) that accounts for variability in genetic sequences within a species, wherein the reference DAG comprises a first path corresponding to a first allele and a second path corresponding to a second allele at a position in the reference DAG, wherein the first allele comprises a genetic structural variation, wherein the plurality

of sequence reads that include portions of the genetic structural variation properly align to the reference DAG; and determine a genotype for the genetic sample based upon the number of sequence reads aligning to the first path and the second path, wherein the alignment of two or more related reads is used to determine the genotype, and the determined genotype comprises the structural variation.

The methods and systems of the invention transform linear, local sequence alignment processes such as, for example, Smith-Waterman-Gotoh, into multi-dimensional alignment algorithms that provide increased parallelization, increased speed, increased accuracy, and the ability to align reads through an entire genome. Algorithms of the invention provide for a "look-

CLAIMS

1. A method of genotyping a genetic sample, comprising:

obtaining a plurality of sequence reads corresponding to a genetic sample, wherein the plurality of sequence reads include portions of a structural variation;

aligning the plurality of sequence reads to a reference directed acyclic graph (DAG) that accounts for variability in genetic sequences within a species, the reference DAG comprising a first path corresponding to a first allele and a second path corresponding to a second allele at a position in the reference DAG, wherein the first allele comprises a genetic structural variation, wherein the plurality of sequence reads that include portions of the genetic structural variation properly align to the reference DAG; and

determining a genotype for the genetic sample based upon the number of sequence reads aligning to the first path and the second path, wherein the alignment of two or more related reads is used to determine the genotype, and the determined genotype comprises the structural variation.

2. The method of claim 1, wherein the method does not include comparing the sequence reads to a variant call format (VCF) file or a single nucleotide polymorphism database (dbSNP).

3. The method of claim 1, wherein the reference DAG further comprises a third path corresponding to a third allele at the position in the reference DAG.

4. The method of claim 1, further comprising determining an allele frequency for the sample based upon the number of sequence reads that align to the first path and the second path .

5. The method of claim 1, further comprising determining a confidence value for the genotype based upon the number of overlapping base pairs between a sequence read and one of the first path and the second path, wherein more overlap correlates with greater confidence.

6. The method of claim 1, wherein the reference DAG represents a chromosome.

7. The method of claim 1, wherein the reference DAG represents a genome.

8. The method of claim 1, wherein the first path and the second path differ from each other by a base deletion, a base insertion, or a polymorphism.

9. The method of claim 1, wherein the sequence reads are obtained by sequencing a genetic sample of a subject with a method selected from Sanger sequencing, pyrosequencing, ion semiconductor sequencing, sequencing by synthesis, sequencing by ligation, and single-molecule real-time sequencing.

10. The method of claim 1, wherein the reference DAG construct comprises greater than about 1,000 base pairs.

11. The method of clam 10, wherein the reference DAG construct comprises greater than about 1,000,000 symbols.

12. The method of claim 1, wherein the plurality of sequence reads comprises greater than about 1000 sequence reads.

13. The method of claim 1, wherein the majority of the plurality of sequence reads are greater than about 50 base pairs in length.

14. The method of claim 13, wherein the majority of the plurality of sequence reads are greater than about 100 base pairs in length.

15. The method of claim 1, wherein the genetic sample originates in blood, urine, saliva, sputum, stool, nipple aspirate, sweat, hair follicle, buccal swab, or tissue.

16. The method of claim 15, further comprising isolating a plurality of nucleic acids from the sample.

17. The method of claim 1, further comprising diagnosing a disease based upon the genotype of the sample.

18. A computer-implemented method of genotyping a genetic sample, comprising:

providing a plurality of sequence reads corresponding to a genetic sample, wherein the plurality of sequence reads include portions of a structural variation;

aligning the plurality of sequence reads to a reference directed acyclic graph (DAG) that accounts for variability in genetic sequences within a species, the reference DAG comprising nodes, wherein the reference DAG comprises a first path corresponding to a first allele and a second path corresponding to a second allele at a position in the reference DAG using a computer processor, wherein the first allele comprises a genetic structural variation, wherein the plurality of sequence reads that include portions of the genetic structural variation properly align to the reference DAG; and

determining a genotype for the genetic sample based upon the number of sequence reads aligning to the first path and the second path, wherein the alignment of two or more related reads is used to determine the genotype, and the determined genotype comprises the structural variation.

19. The method of claim 18, wherein the plurality of sequence reads are provided in a non-transitory computer-readable medium.

20. The method of claim 18, wherein the reference DAG is provided in in a non-transitory computer-readable medium.

21. A system for aligning a plurality of reads comprising a processor and memory, wherein the memory comprises instructions that, when executed, cause the processor to:

obtain a plurality of sequence reads corresponding to a genetic sample, wherein the plurality of sequence reads include portions of a structural variation;

align the plurality of sequence reads to a reference directed acyclic graph (DAG) that accounts for variability in genetic sequences within a species, wherein the reference DAG comprises a first path corresponding to a first allele and a second path corresponding to a second allele at a position in the reference DAG, wherein the first allele comprises a genetic structural variation, wherein the plurality of sequence reads that include portions of the genetic structural variation properly align to the reference DAG; and

determine a genotype for the genetic sample based upon the number of sequence reads aligning to the first path and the second path, wherein the alignment of two or more related reads is used to determine the genotype, and the determined genotype comprises the structural variation.

22. The system of claim 21, wherein aligning comprises:

simultaneously comparing a string of symbols corresponding to a sequence read to a plurality of different genotypes at the position,

scoring overlaps between the string of symbols and each of the first path and the second path, wherein a higher score corresponds to a greater amount of overlap; and

identifying the overlaps corresponding to the highest score for the sequence read.

23. The system of claim 22, further comprising writing a file to memory corresponding to the identified genotype.

24. The system of claim 21, wherein the plurality of sequence reads are provided in a non-transitory computer-readable medium.

25. The system of claim 21, wherein the reference DAG is provided in in a non-transitory computer-readable medium.

26. The system of claim 21, wherein the system comprises a plurality of processors, and wherein each processor is configured to align a portion of the plurality of sequence reads to the reference DAG.

27. The system of claim 21, wherein the instructions additionally cause the processor to output the determined genotype for the sample.