yet a different genotype. In other instances, the sequences of interest have moved (relative to the "normal" position) because of their proximity to a structural variation.

Incorporating structural variations into the state-of-the art genotyping methods is very difficult, however, because there are many known variants from which any particular read might originate. For each N structural variations, there are approximately 2^N different references that have to be compared to the assembled sequence in order to genotype the sequence. In other words, to accommodate one structural variant, the assembled sequence must be compared to at least 2 separate reference sequences to genotype the sample, but accommodating 20 possible structural variations requires comparing the sequence to roughly one million different references. Even using state-of—the-art methods with parallel computing, this is a very expensive proposition. Furthermore, this combinatorial explosion makes it impossible to feasibly genotype longer sequences that include hundreds of possible structural variations. Thus, approximations are made in the name of reducing computational time. In other words, current methods do not fairly represent the structural variations that are commonly found in many genomes.

SUMMARY

The ~~invention~~ disclosure provides methods and systems for efficiently genotyping sequence reads by aligning the reads directly to a reference sequence construct that simultaneously accounts for multiple alleles at multiple loci in the genome of the organism. Furthermore, the methods and systems of the ~~invention~~ disclosure make it possible to deal with structural variations in an efficient way, greatly reducing the computing power necessary to genotype genetic samples using next generation sequencing (N.G.S.). Additionally, because the reference sequence construct accounts for the various possible alleles within the construct, it is possible to directly genotype the sample by merely aligning the reads of the sample to the construct. Particular patterns of alignment are only possible for particular genotypes, thus it is not necessary to compare the assembled sequence to a reference sequence and then compare the variations to mutation files associated with that reference. 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; 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; and genotyping the sample based upon the alignment of the plurality of

sequence reads against the reference sequence construct. In another 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; align the plurality of sequence reads to a reference sequence construct that represents different genotypes as alternative sequences at a first position in the construct; and genotype the sample based upon the alignment of the plurality of sequence reads against the reference sequence construct. In yet 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; 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.

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-