CLAIMS

1. A non-transitory computer-readable medium comprising instructions that, when executed by at least one processor, cause the at least one processor to:

identify candidate structural variants that satisfy a threshold quantity of occurrences within a genomic sample database;

select, from the candidate structural variants, structural variant haplotypes comprising one or more structural variants in phase with flanking variants within nucleotide sequences of the genomic sample database;

identify, from a linear reference genome, reference haplotypes corresponding to the structural variant haplotypes; and

generate, for mapping and aligning nucleotide reads of a genomic sample, a structural variation graph genome comprising alternate contiguous sequences representing the structural variant haplotypes and reference sequences representing the reference haplotypes.

2. The non-transitory computer-readable medium of claim 1, further comprising instructions that, when executed by the at least one processor, cause the at least one processor, to select the structural variant haplotypes by selecting, from the candidate structural variants, particular structural variant haplotypes that satisfy an additional threshold quantity of occurrences at particular genomic regions.

3. The non-transitory computer-readable medium of claim 2, further comprising instructions that, when executed by the at least one processor, cause the at least one processor to select the structural variant haplotypes by selecting particular structural variant haplotypes that are in phase with particular flanking variants within contiguous sequences of the genomic sample database.

4. The non-transitory computer-readable medium of claim 3, further comprising instructions that, when executed by the at least one processor, cause the at least one processor to select the particular structural variant haplotypes by:

selecting a first structural variant haplotype in phase with a first flanking variant within a first contiguous sequence of the genomic sample database; and

selecting a second structural variant haplotype in phase with a second flanking variant within a second contiguous sequence of the genomic sample database.

5. The non-transitory computer-readable medium of claim 3, further comprising instructions that, when executed by the at least one processor, cause the at least one processor to generate the structural variation graph genome comprising particular alternate contiguous sequences representing the particular structural variant haplotypes and the particular flanking variants.

6. The non-transitory computer-readable medium of any one of claims 1-5, wherein a flanking variant comprises a single nucleotide polymorphism (SNP), a deletion of less than fifty base pairs, or an insertion of less than fifty base pairs.

7. The non-transitory computer-readable medium of any one of claims 1-5, further comprising instructions that, when executed by the at least one processor, cause the at least one processor to:

identify, from the genomic sample database, alternate haplotypes comprising one or more of a single nucleotide polymorphism (SNP), a deletion of less than fifty base pairs, or an insertion of less than fifty base pairs; and

generate the structural variation graph genome further comprising alternate nucleobases or additional alternate contiguous sequences representing the alternate haplotypes.

8. The non-transitory computer-readable medium of any one of claims 1-5, further comprising instructions that, when executed by the at least one processor, cause the at least one processor to identify the candidate structural variants by selecting structural variants representing one or more of a deletion of more than fifty base pairs, an insertion of more than fifty base pairs, a duplication of more than fifty base pairs, an inversion, a translocation, or a copy number variation (CNV).

9. The non-transitory computer-readable medium of any one of claims 1-5, wherein the at least one processor comprises a configurable processor and executing the at least one processor comprises configuring the configurable processor.

10. A system comprising:

at least one processor; and

a non-transitory computer readable medium comprising instructions that, when executed by the at least one processor, cause the system to:

identify candidate structural variants that satisfy a threshold quantity of occurrences within a genomic sample database;

select, from the candidate structural variants, structural variant haplotypes comprising one or more structural variants in phase with flanking variants within nucleotide sequences of the genomic sample database;

identify, from a linear reference genome, reference haplotypes corresponding to the structural variant haplotypes; and

generate, for mapping and aligning nucleotide reads of a genomic sample, a structural variation graph genome comprising alternate contiguous sequences representing the structural variant haplotypes and reference sequences representing the reference haplotypes.

11. The system of claim 10, further comprising instructions that, when executed by the at least one processor, cause the system to select the structural variant haplotypes by:

selecting, from the candidate structural variants, a first structural variant haplotype that satisfies an additional threshold quantity of occurrences at a first genomic region; and

selecting, from the candidate structural variants, a second structural variant haplotype that satisfies the additional threshold quantity of occurrences at a second genomic region.

12. The system of claim 10 or 11, further comprising instructions that, when executed by the at least one processor, cause the system to select the structural variant haplotypes by:

selecting a first structural variant haplotype in phase with a first flanking variant within a first contiguous sequence of the genomic sample database; and

selecting a second structural variant haplotype in phase with a second flanking variant within a second contiguous sequence of the genomic sample database.

13. The system of claim 12, wherein the first flanking variant or the second flanking variant comprises a single nucleotide polymorphism (SNP), a deletion of less than a threshold number of base pairs, or an insertion of less than the threshold number of base pairs.

14. The system of claim 12, further comprising instructions that, when executed by the at least one processor, cause the system to identify the candidate structural variants by selecting structural variants representing one or more of a deletion of more than a threshold number of base pairs, an insertion of more than the threshold number of base pairs, a duplication of more than the threshold number of base pairs, an inversion, a translocation, or a copy number variation (CNV).

15. The system of claim 12, further comprising instructions that, when executed by the at least one processor, cause the system to generate the structural variation graph genome comprising:

a first alternate contiguous sequence representing a first structural variant haplotype and a first flanking variant; and

a second alternate contiguous sequence representing a second structural variant haplotype and a second flanking variant.

16. The system of claim 10 or 11, further comprising instructions that, when executed by the at least one processor, cause the system to:

generate an alignment file that maps the structural variant haplotypes to genomic coordinates of the reference haplotypes within the linear reference genome; and

generate the structural variation graph genome by associating, within an organization structure, the alternate contiguous sequences representing the structural variant haplotypes with identifiers for the genomic coordinates of the reference haplotypes.

17. The system of claim 16, further comprising instructions that, when executed by the at least one processor, cause the system to:

generate the alignment file by generating a Sequence Alignment/Map (SAM) liftover file that maps the structural variant haplotypes to the genomic coordinates of the reference haplotypes; and

generate the structural variation graph genome utilizing the organization structure by associating, within a hash table, nucleobase identifiers for nucleobases from the alternate contiguous sequences with values representing the genomic coordinates of the reference haplotypes.

18. The system of claim 10 or 11, further comprising instructions that, when executed by the at least one processor, cause the system to generate the structural variation graph genome by ordering a subset of alternate contiguous sequences corresponding to a genomic region according to frequency within the genomic sample database.

19. A computer-implemented method comprising:

identifying candidate structural variants that satisfy a threshold quantity of occurrences within a genomic sample database;

selecting, from the candidate structural variants, structural variant haplotypes comprising one or more structural variants in phase with flanking variants within nucleotide sequences of the genomic sample database;

identifying, from a linear reference genome, reference haplotypes corresponding to the structural variant haplotypes; and

generating, for mapping and aligning nucleotide reads of a genomic sample, a structural variation graph genome comprising alternate contiguous sequences representing the structural variant haplotypes and reference sequences representing the reference haplotypes.

20. The computer-implemented method of claim 19, wherein selecting the structural variant haplotypes comprises selecting, from the candidate structural variants, particular structural variant haplotypes that satisfy an additional threshold quantity of occurrences at particular genomic regions.