1. What part of a template, when sequenced, is represented as a read?

a. Fragment c. Index

b. Adapter d. Insert

2. Reads are stored in ASCII text files.

a. True

b. False

3. Reads are organized into records.

a. True

b. False

4. Which type of text file stores read records?

a. FASTA **c. FASTQ**

b. QUALS d. CONTIG

5. Read records contain (check all that apply)

a. Header line **c. Nucleotide sequence line**

b. Quals header line **d. Quals sequence line**

6. Records in FASTQ files are spaceless.

a. True

b. False

7. Read nucleotide sequences within a record are wrapped at 80 characters.

a. True

b. False

8. The ideal number of lines per record in a FASTQ file is _____.

a. 2 c. 3

b. 4 d. 5

9. FASTQ is actually a combination of two files. These two files are the...

a. FASTA and FASTAP **c. FASTA and QUALS**

b. QUALS and FASTAP d. QUALS and FASTAN

10. What two types of genome reconstruction algorithms are used to reconstruct the genome from reads or contigs found within the FASTQ or FASTA files?

a. Assembly and Construction c. String comparison and suffix tree

b. Assembly and alignment d. Pileup and string matching

11. What is the algorithmic complexity of assembly algorithms?

a. O(n^2) c. O(nm)

b. O(n-m) d. O(n^m)

12. One of the biggest problems with assembly algorithms is that they use up too much memory?

a. True

b. False

13. Alignment is less complex than assembly.

a. True

b. False

14. Alignment algorithms have a linear complexity of O(nm).

a. True

b. False

15. Which alignment algorithms do we generally use for whole genome alignment on SBPLA?

a. Smith-Waterman aligner c. Needleman-Wunsch

b. FASTAP **d. Burrows Wheeler Aligner**

16. The Burrows-Wheeler Aligner uses a BW Transform as a searchable structure for efficient substring matching.

a. True

b. False

17. The Smith-Waterman aligner is a local aligner.

a. True

b. False

18. All non Exact Match aligners are based on a scoring scheme for determining the best alignments.

a. True

b. False

19. Assembly algorithms use the reference genome.

a. True

b. False

20. The reference genome is used to make the BWT.

a. True

b. False

21. Smith-Waterman aligners are local aligners.

a. True

b. False

22. CIGAR strings are found in FASTQ files.

a. True

b. False

23. BAM files are _____ SAM files.

a. Unix

b. Binary

c. UTF

24. BAM stands for _____.

a. Biological Analaysis Map

b. Binary Sequence Alignment Map

c. Big Array Message

25. CIGAR string are important for which variant calling data preprocessing step?

a. Base recalibration

b. Realignment around indels

c. Variant recalibration

26. BAM files can be _____ or _____ sorted.

a. Coordinate

b. Queryname

c. Sequence

27. The main columns in the BAM format are.....

a. Reference chromosome

b. Mapping position

c. Annotations

28. BAM files must be _____ sorted in order to create a BAI file for them.

a. Queryname

b. Coordinate

c. Lexicographically

29. Two main institutions that curate reference genomes are _____ and _____?

UCSC - University of California Santa Cruz

GCR - Genome Reference Consortium

30. The main difference between GRCh and Hg reference genomes is in _____.

a. contig nomenclature

b. contig lengths

c. missing headers

31. FASTA index files are used to _____ access FASTA files.

a. randomly

b. iteratively

c. sequentially

32. FAI files have exact lengths of contigs in terms of nucleotides.

a. True

b. False

33. BED files are _____ files.

a. interval files

b. annotation files

c. both

34. Two step aligners have which two steps?

a. seed

b. extend

c. elongate

35. What does this CIGAR string mean: ? 1S10M10D3H

1 Soft Clipping, 10 Alignment Matches (can be a sequence match or mismatch), 10 Deletion and 3 Hard Clippings

36. How large are BAM files typically?

a. 100GB - 500GB

b. 10GB - 100GB

c. 1TB

37. BAM files are compressed SAM files.

a. True

b. False

38. BAM Index files or BAI files contain which two types of data?

a. Block number

b. Aligment line number within the block

c. Sort number

39. BAM file reference sequences begin with 0 while SAM file reference seqs begin with 1.

a. True

b. False

40. The BWA-MEM is the aligner used to produce SAM files on the SBPLA.

a. True

b. False