Basic Bioinformatics and Long-Read Sequencing Data Analysis

Final Test   Name: _____Date: _____

Instructions: Answer all questions. Write commands exactly as you would type them.

Part 1: Basic Shell Commands   1. What command would you use to find out which directory you are currently in?

+Answer: _____

2. You are in **/home/user1/** and want to navigate to **/home/user1/seq-analysis/results/**. Write the command:

+Answer: _____

3. Match the following commands with their functions: - `mkdir` _____ a) Copy files - `cp` _____ b) List directory contents
- `mv` _____ c) Create directories - `ls` _____ d) Move/rename files

4. What does the `>` symbol do in this command?

```
wc -l *.pdb > lengths.txt
```

+Answer: _____

5. What is the purpose of the pipe symbol | in shell commands? Give an example.

+Answer: _____

_____

Part 2: File Formats & Quality Control 6. In a FASTQ file, what does each of the 4 lines represent for a single sequence? - Line 1: _____ - Line 2: _____
- Line 3: _____ - Line 4: _____

7. To view the contents of a compressed file `sample.fastq.gz` without uncompressing it, which command would you use?

+Answer: _____

_____

Part 3: Sequence Analysis Workflow   8. Put these file formats in the correct order for a variant calling workflow: _____ VCF _____ FASTQ _____ SAM _____ BAM

9. Why must we index the reference genome before alignment?

+Answer: _____

10. What is the main difference between SAM and BAM file formats?

+Answer: _____

11. After variant calling with freebayes, you have 500 variants. You filter with:

```
bcftools filter -i 'QUAL>20' input.vcf > output.vcf
```

What does this command do and why is it important?

+Answer: _____

Bonus Question What does SSH stand for and why do bioinformaticians need to use it?

+Answer: _____