INTRODUCTORY POPULATION GENOMICS EXAM

Total possible marks: 80

Please answer all questions without using the internet or your computer.

SECTION 1 – Linux and the command line (10 marks)

1. In the Linux shell, state which command would you use to:
      i) Change directory:
      ii) Copy a file:
      iii) Delete a folder:

(3 marks)

2. Describe what these symbols might be used for in the command line:
      i) |
      ii) >
      iii) ..

(3 marks)

3. Write two or three sentences on why we need to use Linux and the command line for population genomic analyses"

(4 marks)

SECTION 2 – File formats (10 marks)

4. Which of these file types are commonly used as inputs for plink (place an X by ALL correct answers):
      i) .bed
      ii) .env
      iii) .fam
      iv) .bam
      v) .ped
      vi) .bim
      vii) .map
      viii) .fasta
      ix) .genome

(5 marks)

5. Which of these describes how data are stored in a .vcf file (place an X by ONE correct answer):

     i) individuals in rows, variants in columns
     ii) individuals in rows, genes in columns
     iii) variants in rows, individuals in columns
     iv) genes in rows, individuals in columns

(1 mark)

6. Write two or three sentences on why we might store large-scale genome data in vcf format, rather than as whole genome .fasta files for each individual?

(4 marks)

SECTION 3 – R (10 marks)

7. In R, which function would you use to:
     i) Read in a comma separated values (.csv) file:
     ii) Calculate the mean of a column:
     iii) Change the column names of a data frame:
     iv) Combine two data frames (one on top of the other):

(4 marks)

8. Which of these are valid ways to extract the column "species" from the data frame "dd":
     i) dd^species
     ii) dd[,"species"]
     iii) dd$species
     iv) dd$SPECIES

(2 marks)

9. List four things that R is used for:
     i)
     ii)
     iii)
     iv)

(4 marks)

SECTION 4 – Genetic diversity (10 marks)

10. What can we use to estimate the expected genotype frequencies of an equilibrium population (place an X next to ONE correct answer):
      i) Darwin's theory of evolution by natural selection
      ii) Mendel's law of particulate inheritance
      iii) The Hardy-Weinberg equation

(1 mark)

11. Which of these are measures of genetic diversity within a population (place an X next to ALL correct answers):
      i) Nucleotide diversity
      ii) Expected heterozygosity
      iii) EigenGWAS
      iv) Observed heterozygosity
      v) Proportion polymorphic loci (P)
      vi) VCF

(4 marks)

12. Write two or three sentences on how genomics can help us to study genetic diversity within populations:

(5 marks)

SECTION 5 – Population structure (10 marks)

13. Population structure causes deviations from Hardy-Weinberg equilibrium because of what effect (place an X next to ONE correct answer):
      i) The butterfly effect
      ii) The Wahlund effect
      iii) The placebo effect

(1 mark)

14. Which of these approaches are valid methods/statistics for looking at population structure (placed an X next to ALL correct answers):
      i) Linkage mapping
      ii) F_{ST}
      iii) dN/dS
      iv) PCA
      v) Admixture
      vi) STRUCTURE

(4 marks)

15. Write two or three sentences on how genomics can help us to study population structure:

(5 marks)

SECTION 6 – Hybridisation and introgression (10 marks)

16. What is a hybrid (place an X next to ONE correct answer)?
       i) An individual from an outbred population
       ii) An offspring that has undergone natural selection
       iii) An offspring resulting from the crossing of two species

(1 mark)

17. Which of these programs/approaches can be used to identify hybridisation and/or introgression (place a circle next to ALL correct answers):
       i) EigenGWAS
       ii) HiEST
       iii) Admixture
       iv) Bayenv
       v) PCA
       vi) ABBA-BABA tests

(4 marks)

18. Write two or three sentences on the difference between hybridisation and introgression

(5 marks)

SECTION 7 – Natural selection (10 marks)

19. Which of these statements is **false** (place an X next to ONE answer):
       i) Natural selection can result in a change in allele frequencies within populations
       ii) Natural selection and evolution are the same thing
       iii) Natural selection can result in phenotypic evolution

(1 mark)

20. Which of these programs/statistics can be used to identify regions of the genome subject to natural selection (place an X next to ALL correct answers)

i) EigenGWAS
ii) nano
iii) Admixture
iv) Tajima's D
v) FST outlier analysis
vi) Nucleotide diversity

(4 marks)

21. Write two or three sentences on how genomics can help us to understand natural selection

(5 marks)

SECTION 8 – Landscape and seascape genomics (10 marks)

22. Which of these statements are **true** (place an X next to ALL correct answers)

i) Environmental association analyses can be confounded by population structure
ii) Environmental variation can result in divergent selection
iii) Environmental association analyses can't be carried out in marine organisms

(2 marks)

23. Describe what the software BayEnv does:

(3 marks)

24. Write two or three sentences on how landscape genomics is different from landscape genetics

(5 marks)

END