1

THE EFFECTS OF ALIGNMENT ERROR AND ALIGNMENT FILTERING ON THE SITEWISE DETECTION OF POSITIVE SELECTION

1.1 INTRODUCTION

1.1.1 *Methods for detecting sitewise positive selection*

1.1.2 *Substitution and indel processes in simulating protein-coding sequence evolution*

1.2 MODELS AND PARAMETERS FOR SIMULATING THE EVOLUTION OF MAMMALIAN GENES

1.2.1 *Distribution of selective pressures*

1.2.2 *Phylogenetic tree size and shape*

1.2.3 *Frequency and size distribution of insertions and deletions*

1.3 ANALYSIS OF THE ALIGNMENT ERROR SIMULATION RESULTS

1.4 METHODS FOR FILTERING ALIGNMENTS

1.5 ANALYSIS OF THE ALIGNMENT FILTERING SIMULATION RESULTS

2

THE EFFECTS OF ALIGNMENT ERROR AND ALIGNMENT FILTERING ON DETECTING POSITIVE SELECTION IN GENES

2.1 INTRODUCTION

2.1.1 *Existing methods for detecting gene-wide positive selection*

2.2 THE APPLICATION OF SITEWISE ESTIMATES TO THE GENEWISE DETECTION OF POSITIVE SELECTION

2.3 ANALYSIS OF THE GENEWISE DETECTION RESULTS

2.4 CONCLUSIONS AND FURTHER WORK

3

PATTERNS OF SITEWISE SELECTION IN MAMMALIAN PROTEIN-CODING GENES

3.1 INTRODUCTION

3.1.1 *The Mammalian Genome Project*

3.1.2 *The Sitewise Likelihood Ratio test*

3.1.3 *Data quality concerns: alignment and sequencing error*

3.1.4 *Gene trees, genomic alignments, and low-coverage genomes in the Ensembl database*

3.2 METHODS TO IDENTIFY ORTHOLOGOUS SUBTREES WITHIN LARGE MAMMALIAN GENE FAMILIES

3.3 ANALYSIS OF THE GENOME-WIDE SET OF ORTHOLOGOUS MAMMALIAN TREES

3.4 ANALYSIS OF THE GLOBAL DISTRIBUTION OF MAMMALIAN SELECTIVE PRESSURES

3.5 ANALYSIS OF SITEWISE ESTIMATES FROM THREE MAMMALIAN SUB-CLADES

3.6 EVALUATION OF THE EFFECT OF GC CONTENT, RECOMBINATION RATE, AND CODON USAGE ON SITEWISE DNDS ESTIMATES AND THE DETECTION OF POSITIVE SELECTION

3.6.1 *Mammalian sitewise selective pressures are not subject to strong effects of biased gene conversion*

3.6.2 *Mammalian sitewise selective pressures suggest increased efficacy of natural selection in regions of high recombination*

3.7 CONCLUSIONS AND FUTURE WORK

4

THE USE OF SITEWISE SELECTIVE PRESSURES TO CHARACTERISE THE EVOLUTION OF GENES AND DOMAINS IN MAMMALS

4.1 COMPARISON OF SITEWISE RESULTS TO PREVIOUSLY DESCRIBED SETS OF POSITIVELY SELECTED GENES

4.2 USING SITEWISE SELECTIVE PRESSURES TO CHARACTERISE THE EVOLUTION OF GENES

4.2.1 *Identifying genes subject to positive selection*

4.2.2 *Identifying genes subject to strong or weak purifying selection*

4.3 USING SITEWISE SELECTIVE PRESSURES TO CHARACTERISE THE EVOLUTION OF PROTEIN DOMAINS

4.3.1 *Identifying protein domains subject to positive selection*

4.3.2 *Identifying protein domains subject to strong or weak purifying selection*

5

THE EVOLUTION OF PROTEIN-CODING GENES IN GORILLA AND THE AFRICAN APES

5.1 INTRODUCTION

5.1.1 *The gorilla and other primate genome projects*

5.1.2 *Incomplete lineage sorting*

5.1.3 *Effective population sizes of extant and ancestral primate populations*

5.1.4 *Measuring shifts in selective pressures using branch-specific likelihood ratio tests*

5.1.5 *Data quality concerns: sequencing, assembly and alignment error*

5.2 CONSTRUCTING CODON ALIGNMENTS OF ONE-TO-ONE ORTHOLOGOUS GENES IN SIX PRIMATE SPECIES

5.2.1 *Identification of genes with one-to-one homology*

5.2.2 *Collection of homologous DNA sequences from genome- or transcript-based multiple alignments*

5.2.3 *Filtering sequence regions with low sequence quality*

5.2.4 *Filtering sequence regions with high substitution counts*

5.2.5 *Filtering sequence regions with evidence of incomplete lineage sorting*

5.3 ANALYSIS OF PATTERNS OF DUPLICATION AND DELETION IN PRIMATE GENE FAMILIES

5.4 ANALYSIS OF THE LIKELIHOOD RATIO TEST RESULTS

5.4.1 *Genes with evidence for acceleration and deceleration in the human, chimpanzee and gorilla terminal lineages*

5.4.2 *Genes with evidence for acceleration in the African great ape ancestral branch*

5.4.3 *Genes with evidence for positive selection based on the branch-site test*

6

GORILLA PART 2

6.1 ANALYSIS OF INCOMPLETE LINEAGE SORTING IN THE AFRICAN GREAT APES WITHIN AND NEARBY PROTEIN-CODING GENES

6.2 ANALYSIS OF DN/DS LEVELS IN SIX PRIMATE GENOMES

6.2.1 *Genome-wide dN/dS in six primates and their ancestors*

6.2.2 *Genome-wide dN/dS in regions of differing sitewise constraint*

6.2.3 *Analysis of the impact of sequence and alignment filtering on primate dN/dS estimates*

6.3 CONCLUSIONS AND FUTURE WORK