SLiM

Workshop Series

#21: Nucleotide-based Models

Benjamin C. Haller   Messer Lab, Cornell University

Nucleotide-based Models

- Tracks the nucleotide sequence of every genome
- Mutations have an associated nucleotide
- Mutation rates are sequence-dependent
- Realistic gene conversion, including gBGC

```
GAATGTCGGTTAGAGCAACCTAGCTTCTCAGATCGCAATA
GAATGTCCGTTAGAGCAACCTAGCTTCTCAGATGGCTATA
GAATGTCCGTTAGAGCAACCTAGCTTCTCAGATGGCCATA
GAATGTCGGTTAGAGCATCCTAGCTTCTCAGATCGCAATA
GAATGTCGGTTAGAGCAACCTAGCTTCTCAGATCGCAATA
GAATGTCCGTTAGAGCAACCTAGCTTCTCAGATGGCAATA
GAATGTCGGTTAGAGCATCCTAGCCTCTCAGATGGCAATA
GAATGTCGGTTAGAGCATCCTAGCTTCTCAGATCGCAATA
```

Nucleotide-based Models

- Are *declared* as nucleotide-based
  - an option on `initializeSLiMOptions()`

- Have an *ancestral sequence* (reference sequence)
  - set with `initializeAncestralNucleotides()`

- Use nucleotide-based *mutation types*
  - set up with `initializeMutationTypeNuc()`
  - this establishes an `"l"` stacking policy

- Define sequence-based *mutation rate matrix*
  - passed to `initializeGenomicElementType()`

Ancestral sequence updating

- The ancestral sequence is the reference
  - used whenever a mutation is not present

- It updates whenever a mutation fixes

Ancestral sequence updating

- The ancestral sequence is the reference
  - used whenever a mutation is not present

- It updates whenever a mutation fixes

GAATG

GAATG
GAATG
GAATG
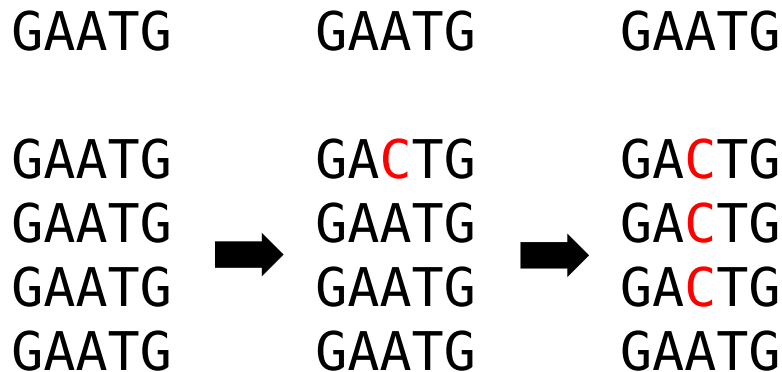GAATG

Ancestral sequence updating

- The ancestral sequence is the reference
    - used whenever a mutation is not present

- It updates whenever a mutation fixes

```
GAATG        GAATG


GAATG        GACTG
GAATG   ➡    GAATG
GAATG        GAATG
GAATG        GAATG
```

Ancestral sequence updating

- The ancestral sequence is the reference
  - used whenever a mutation is not present

- It updates whenever a mutation fixes

```
GAATG          GAATG          GAATG


GAATG          GACTG          GACTG
GAATG    →     GAATG    →     GACTG
GAATG          GAATG          GACTG
GAATG          GAATG          GAATG
```
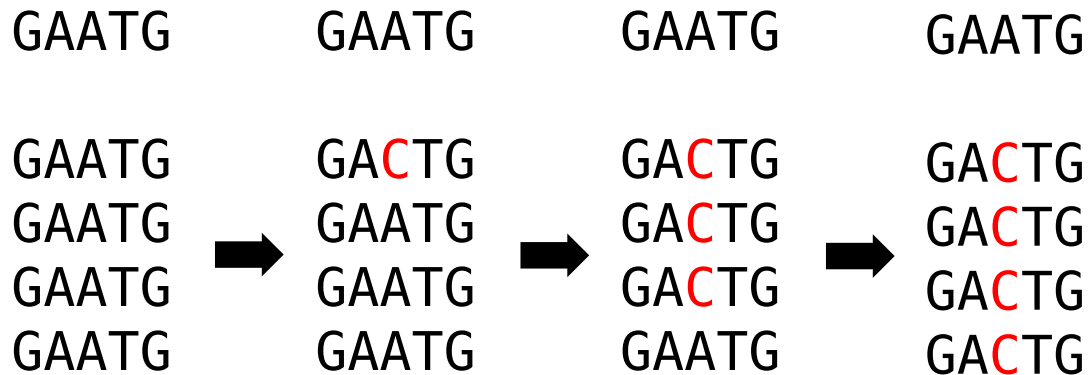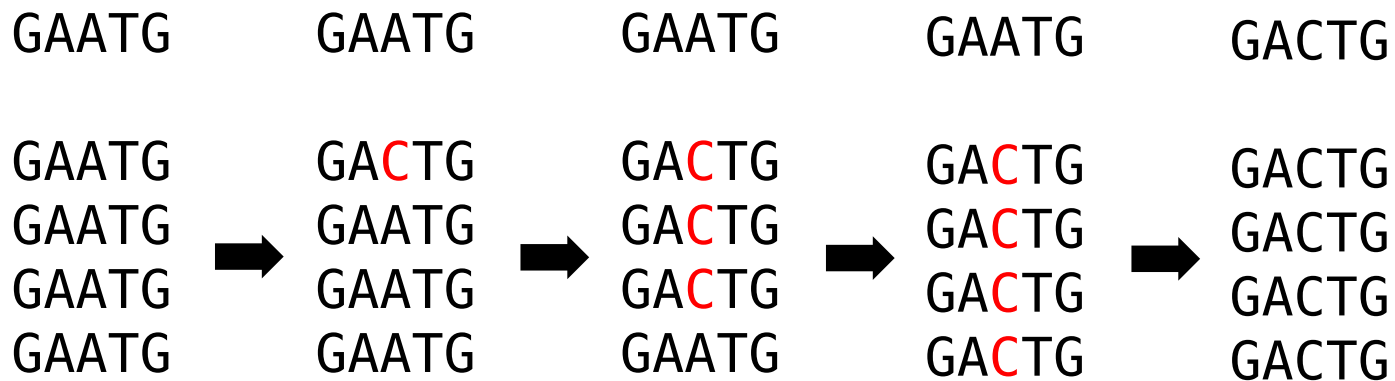
Ancestral sequence updating

- The ancestral sequence is the reference
  - used whenever a mutation is not present
- It updates whenever a mutation fixes

```
GAATG        GAATG        GAATG        GAATG

GAATG        GACTG        GACTG        GACTG
GAATG   →    GAATG   →    GACTG   →    GACTG
GAATG        GAATG        GACTG        GACTG
GAATG        GAATG        GAATG        GACTG
```

Ancestral sequence updating

- The ancestral sequence is the reference
  - used whenever a mutation is not present
- It updates whenever a mutation fixes

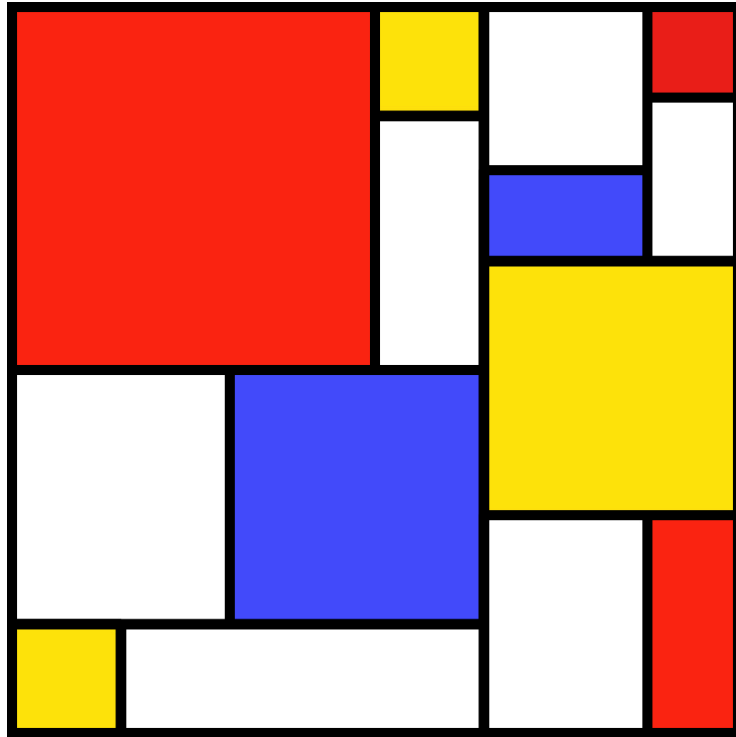|				
GAATG	GAATG	GAATG	GAATG	GACTG
GAATG	GACTG	GACTG	GACTG	GACTG
GAATG	GAATG	GACTG	GACTG	GACTG
GAATG	GAATG	GACTG	GACTG	GACTG
GAATG	GAATG	GAATG	GACTG	GACTG

Nucleotide-based Models

- Optional facilities
  - Trinucleotide-based mutation rates
  - Reading and writing FASTA, VCF files
  - Getting the nucleotide sequence
  - Getting the codon sequence
  - Getting the amino acid sequence
  - Hotspot maps (variable mutation rate)
  - GC-biased gene conversion (gBGC)

A complete nucleotide-based model

```
initialize() {
    defineConstant("L", 1e6);
    initializeSLiMOptions(nucleotideBased=T);
    initializeAncestralNucleotides(randomNucleotides(L));
    initializeMutationTypeNuc("m1", 0.5, "f", 0.0);
    initializeGenomicElementType("g1", m1, 1.0, mmJukesCantor(1e-7));
    initializeGenomicElement(g1, 0, L-1);
    initializeRecombinationRate(1e-8);
}
1 { sim.addSubpop("p1", 500); }
2000 late() { sim.outputFixedMutations(); }
```

- Declared as nucleotide-based
- Sets the ancestral (reference) sequence
- Uses a nucleotide-based `MutationType`
- Defines a mutation rate matrix

SLiM Workshop Exercise #21