

PHYLIP INPUT Format Formalization

June 15, 2007

1 Sequence Data

% n is number of seqs and m is length of sequence

MolData → *Integer Integer MolData*($\mathbb{R}_1, \mathbb{R}_2$)

MolData(n, m) → *MolData*(*dna*, n, m)
| *MolData*(*protein*, n, m)

MolData(*type*, n, m) → *MolSeqData*(*type*, n, m)
| *MolInterData*(*type*, n, m)

% sequence on a single line

MolSeqData(*type*, 0, m) → ϵ

MolSeqData(*type*, n, m) → $\{n > 0\}$ *MolSeqLine*(*type*, m) *newline* *MolSeqData*(*type*, $n-1, m$)

MolSeqLine(*type*, m) → *Name10 MolSeqLineData*(*type*, m)

MolSeqLineData(*type*, 0) → *newline*

MolSeqLineData(*type*, m) → $\{m > 0\}$ *MolSymbol*(*type*) *MolSeqLineData*(*type*, $m-1$)

MolSymbol(*type*) → *MolIgnore MolAcceptable*(*type*)

MolAcceptable(*dna*) → *A | B | C | D | G | H | K | M | N | O | R | S | T | U | V | W | X*

MolAcceptable(*protein*) → *A | B | C | D | E | F | G | H | I | J | K | L | M | N | O | P | Q*

MolIgnore → ϵ | *space MolIgnore | Digit MolIgnore*

% interleaved representation

MolInterData(*type*, n, m) → *MolInterFirstLines*(*type*, n, m) *MolInterRemainingLines*(*type*, n, m)

MolInterFirstLines(*type*, n, m) → *MolInterFirstLine*(*type*, n, m) *MolInterFirstLinesRest*(*type*, n, m)

MolInterFirstLine(*type*, n, m) → *Name10 MolCountData*(*type*, m) $\{\mathbb{R} = \mathbb{R}_2\}$

MolCountData(*type*, m) → $\{m \geq 0\}$ ϵ $\{\mathbb{R} = 0\}$

MolCountData(*type*, m) → $\{m \geq 0\}$ *MolSymbol*(*type*) *MolCountData*(*type*, $m-1$) $\{\mathbb{R} = 0\}$

MolInterFirstLinesRest(*type*, 0, m) → ϵ

MolInterFirstLinesRest(*type*, n, m) → $\{n > 0\}$ *Name10 MolSeqLineData*(*type*, m) *MolInterFirstLinesRest*(*type*, $n-1, m$)

MolInterRemainingLines(*type*, $n, 0$) → ϵ

MolInterRemainingLines(*type*, n, m) → $\{m > 0\}$ *MolInterGroup*(*type*, n, m) *MolInterRemainingLines*(*type*, $n, m-1$)

MolInterGroup(*type*, n, m) → *MolInterFirstGroup*(*type*, m) *newline* *MolInterRestGroup*(*type*, n, m)

MolInterFirstGroup(*type*, m) → *MolCountData*(*type*, m) $\{\mathbb{R} = \mathbb{R}_1\}$

MolInterRestGroup(*type*, 0, m) → ϵ

MolInterRestGroup(*type*, n, m) → *MolSeqData*(*type*, m) *newline* *MolInterRestGroup*(*type*, $n, m-1$)

2 Input Trees

<i>TreeFile</i>	→	<i>Integer</i> <i>newline</i> <i>Trees</i>
<i>Trees</i>	→	<i>Trees</i> (<i>simplebinary</i>)
		<i>Trees</i> (<i>distancebinary</i>)
		<i>Trees</i> (<i>simpletern</i>)
		<i>Trees</i> (<i>disttern</i>)
<i>Trees</i> (<i>type</i>)	→	ϵ
		<i>Tree</i> (<i>type</i>) <i>newline</i> <i>Trees</i> (<i>type</i>)
<i>Tree</i> (<i>simplebinary</i>)	→	<i>Name10</i>
		(<i>Tree</i> (<i>simplebinary</i>) , <i>Tree</i> (<i>simplebinary</i>))
<i>Tree</i> (<i>simpletern</i>)	→	<i>Name10</i>
		(<i>Tree</i> (<i>simpletern</i>) , <i>Tree</i> (<i>simpletern</i>) , <i>Tree</i> (<i>simpletern</i>))
<i>Tree</i> (<i>distancebinary</i>)	→	<i>Name10</i> : <i>RealNumber</i>
		(<i>Tree</i> (<i>distancebinary</i>) , <i>Tree</i> (<i>distancebinary</i>)) : <i>RealNumber</i>
<i>Tree</i> (<i>disttern</i>)	→	<i>Name10</i> : <i>RealNumber</i>
		(<i>Tree</i> (<i>disttern</i>) , <i>Tree</i> (<i>disttern</i>) , <i>Tree</i> (<i>disttern</i>)) : <i>RealNumber</i>

3 Quantitative Characters

<i>QData</i>	→	<i>Integer</i> <i>Integer</i> <i>newline</i> <i>QMatrix</i> { $\mathbb{R}_1, \mathbb{R}_2$ }
	→	<i>Integer</i> <i>Integer</i> <i>newline</i> <i>Alleles</i> (\mathbb{R}_2) <i>newline</i> <i>QMatrix</i> ($\mathbb{R}_1, \mathbb{R}_2, \mathbb{R}_3$)
<i>Alleles</i> (<i>n</i>)	→	<i>Alleles_all</i> (<i>n</i>)
	→	<i>Alleles_noall</i> (<i>n</i>)
<i>Alleles_all</i> (0)	→	ϵ { $\mathbb{R} = 0$ }
<i>Alleles_all</i> (<i>n</i>)	→	<i>Integer</i> <i>Alleles_all</i> (<i>n</i> - 1) { $\mathbb{R} = \mathbb{R}_1 + \mathbb{R}_2$ }
<i>Alleles_noall</i> (0)	→	ϵ $\mathbb{R} = 0$
<i>Alleles_noall</i> (<i>n</i>)	→	<i>Integer</i> <i>Alleles_noall</i> (<i>n</i> - 1) { $\mathbb{R} = \mathbb{R}_2 + \mathbb{R}_k - 1$ }
<i>QMatrix</i> (0, <i>m</i>)	→	ϵ
<i>QMatrix</i> (<i>n</i> , <i>m</i>)	→	<i>Name10</i> <i>QLine</i> (<i>m</i>) <i>newline</i> <i>QMatrix</i> (<i>n</i> - 1, <i>m</i>)
<i>QMatrix</i> (0, <i>m</i> , <i>k</i>)	→	ϵ
<i>QMatrix</i> (<i>n</i> , <i>m</i> , <i>k</i>)	→	<i>Name10</i> <i>QLine</i> (<i>k</i>) <i>newline</i> <i>QMatrix</i> (<i>n</i> - 1, <i>m</i> , <i>k</i>)
<i>QLine</i> (0)	→	ϵ
<i>QLine</i> (<i>m</i>)	→	<i>RealNumber</i> <i>QLine</i> (<i>m</i> - 1)

4 Distance

<i>Distance</i>	→ <i>Integer SimpleMatrix</i> (\mathbb{R}_1)
	<i>Integer LowDiagMatrix</i> ($\mathbb{R}_1, 0$)
<i>SimpleMatrix</i> (0)	→ ϵ
<i>SimpleMatrix</i> (<i>n</i>)	→ <i>Name10 MLine</i> (<i>n</i>) <i>newline SimpleMatrix</i> (<i>n</i> - 1)
<i>MLine</i> (0)	→ ϵ
<i>MLine</i> (<i>n</i>)	→ <i>RealNumber MLine</i> (<i>n</i> - 1)
<i>LowDiagMatrix</i> (0, <i>m</i>)	→ ϵ
<i>LowDiagMatrix</i> (<i>n</i> , <i>m</i>)	→ <i>LDLine</i> (<i>m</i>) <i>newline LowDiagMatrix</i> (<i>n</i> - 1, <i>m</i> + 1)
<i>LDLine</i> (<i>m</i>)	→ <i>Name10 LDLineData</i> (<i>m</i>)
<i>LDLLineData</i> (0)	→ <i>RealNumber</i>
	ϵ
<i>LDLLineData</i> (<i>m</i>)	→ <i>RealNumber LDLLineData</i> (<i>m</i> - 1)