Profile HMMs for sequence families

1. So far we have concentrated on the intrinsic properties of single sequences, such as CpG islands.
2.
3.
4.
5.
6.
7.

Profile HMMs for sequence families

1. So far we have concentrated on the intrinsic properties of single sequences, such as CpG islands.
2. However functional biological sequences typically come in families, so what we are after is identifying the relationship of an individual sequences to a sequence family.
3.
4.
5.
6.
7.

Profile HMMs for sequence families

1. So far we have concentrated on the intrinsic properties of single sequences, such as CpG islands.
2. However functional biological sequences typically come in families, so what we are after is identifying the relationship of an individual sequences to a sequence family.
3. A multiple sequence alignment can show how the sequences in a family relate to each other.
4.
5.
6.
7.

Profile HMMs for sequence families

1. So far we have concentrated on the intrinsic properties of single sequences, such as CpG islands.
2. However functional biological sequences typically come in families, so what we are after is identifying the relationship of an individual sequences to a sequence family.
3. A multiple sequence alignment can show how the sequences in a family relate to each other.
4. Some positions in a multiple sequence alignment are more conserved than others (e.g. helices as opposed to loop regions).
5.
6.
7.

Profile HMMs for sequence families

1. So far we have concentrated on the intrinsic properties of single sequences, such as CpG islands.
2. However functional biological sequences typically come in families, so what we are after is identifying the relationship of an individual sequences to a sequence family.
3. A multiple sequence alignment can show how the sequences in a family relate to each other.
4. Some positions in a multiple sequence alignment are more conserved than others (e.g. helices as opposed to loop regions).
5. Therefore it would be desirable, when identifying new sequence members to concentrate more on conserved features.
6.
7.

Profile HMMs for sequence families

1. So far we have concentrated on the intrinsic properties of single sequences, such as CpG islands.
2. However functional biological sequences typically come in families, so what we are after is identifying the relationship of an individual sequences to a sequence family.
3. A multiple sequence alignment can show how the sequences in a family relate to each other.
4. Some positions in a multiple sequence alignment are more conserved than others (e.g. helices as opposed to loop regions).
5. Therefore it would be desirable, when identifying new sequence members to concentrate more on conserved features.
6. For this task we will discuss a special type of HMM, well suited to modelling multiple alignments; we will call these *profile HMMs*.
7.

Profile HMMs for sequence families

1. So far we have concentrated on the intrinsic properties of single sequences, such as CpG islands.
2. However functional biological sequences typically come in families, so what we are after is identifying the relationship of an individual sequences to a sequence family.
3. A multiple sequence alignment can show how the sequences in a family relate to each other.
4. Some positions in a multiple sequence alignment are more conserved than others (e.g. helices as opposed to loop regions).
5. Therefore it would be desirable, when identifying new sequence members to concentrate more on conserved features.
6. For this task we will discuss a special type of HMM, well suited to modelling multiple alignments; we will call these *profile HMMs*.
7. Profile HMMs are the most popular application of HMMs in molecular biology.

Position specific score matrix

A common feature of protein multiple sequence alignments is that gaps tend to line up with each other, leaving solid blocks with no insertions or deletions.

Position specific score matrix

A common feature of protein multiple sequence alignments is that gaps tend to line up with each other, leaving solid blocks with no insertions or deletions.

A probabilistic model for a conserved region (block) would be to specify independent probabilities $e_i(\alpha)$ of observing amino acid α in position i. The probability of a new sequence x, is:

$$P(x \mid M) = \prod_{i=1}^{L} e_i(x_i)$$

where L is the length of the block.

Position specific score matrix

A common feature of protein multiple sequence alignments is that gaps tend to line up with each other, leaving solid blocks with no insertions or deletions.

A probabilistic model for a conserved region (block) would be to specify independent probabilities $e_i(\alpha)$ of observing amino acid α in position i. The probability of a new sequence x, is:

$$P(x \mid M) = \prod_{i=1}^{L} e_i(x_i)$$

where L is the length of the block. To test for family membership we evaluate, as usual log-odds ratio:

$$S = \sum_{i=1}^{L} \log \frac{e_i(x_i)}{q_{x_i}}$$

Position specific score matrix

A common feature of protein multiple sequence alignments is that gaps tend to line up with each other, leaving solid blocks with no insertions or deletions.

A probabilistic model for a conserved region (block) would be to specify independent probabilities $e_i(\alpha)$ of observing amino acid α in position i. The probability of a new sequence x, is:

$$P(x \mid M) = \prod_{i=1}^{L} e_i(x_i)$$

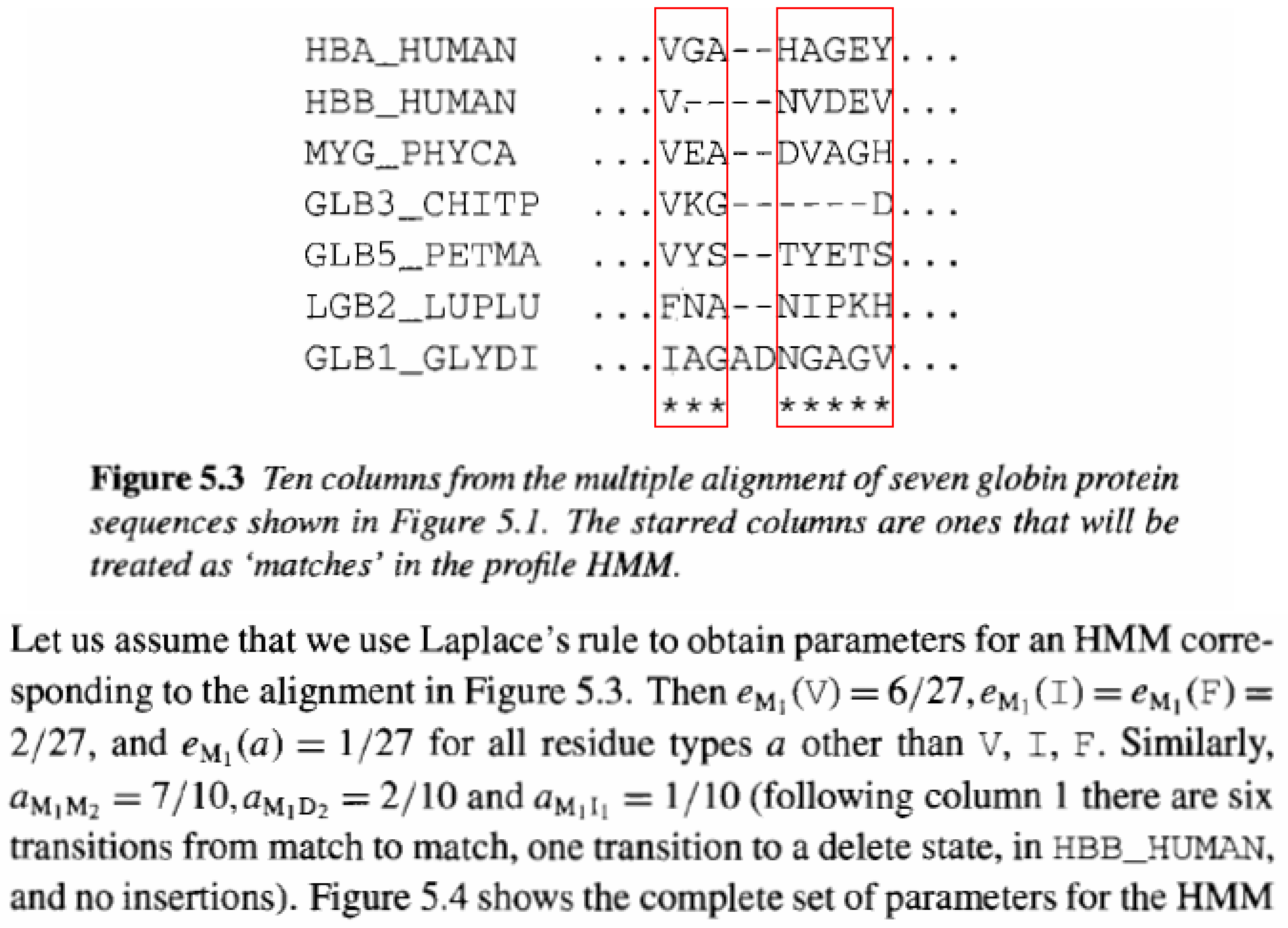
where L is the length of the block. To test for family membership we evaluate, as usual log-odds ratio:

$$S = \sum_{i=1}^{L} \log \frac{e_i(x_i)}{q_{x_i}}$$

The values $\log(e_i(a)/q_a)$ behave like elements in a score matrix $s(a, b)$ where the second index is position i, rather than amino acid b.

Such an approach is known as a *position specific score matrix* (PSSM).

Profile HMMs: αs and es

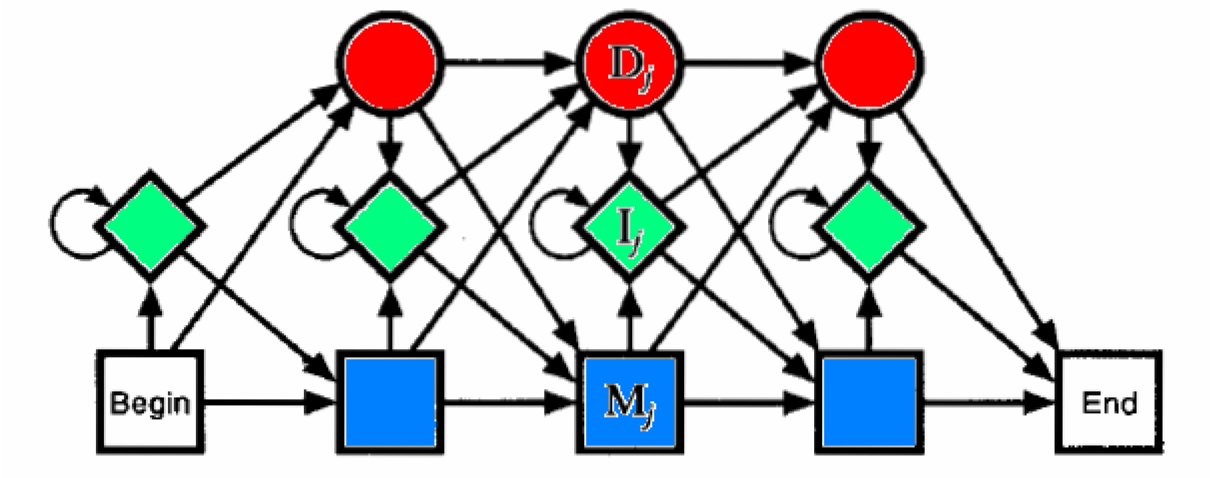


```
HBA_HUMAN    ...VGA--HAGEY...
HBB_HUMAN    ...V----NVDEV...
MYG_PHYCA    ...VEA--DVAGH...
GLB3_CHITP   ...VKG------D...
GLB5_PETMA   ...VYS--TYETS...
LGB2_LUPLU   ...FNA--NIPKH...
GLB1_GLYDI   ...IAGADNGAGV...
                ***  *****
```

Figure 5.3 *Ten columns from the multiple alignment of seven globin protein sequences shown in Figure 5.1. The starred columns are ones that will be treated as 'matches' in the profile HMM.*

Let us assume that we use Laplace's rule to obtain parameters for an HMM corresponding to the alignment in Figure 5.3. Then $e_{M_1}(\mathtt{V}) = 6/27, e_{M_1}(\mathtt{I}) = e_{M_1}(\mathtt{F}) = 2/27$, and $e_{M_1}(a) = 1/27$ for all residue types a other than V, I, F. Similarly, $a_{M_1M_2} = 7/10, a_{M_1D_2} = 2/10$ and $a_{M_1I_1} = 1/10$ (following column 1 there are six transitions from match to match, one transition to a delete state, in HBB_HUMAN, and no insertions). Figure 5.4 shows the complete set of parameters for the HMM

Profile HMMs: transition structure



- **Match**
- **Insertion**
- **Deletion**

Profile HMMs: Viterbi

Let $V_j^M(i)$ be the log-odds score of the best path matching subsequence $x_{1\ldots i}$ to the submodel up to state j, ending with x_i being emitted by state M^j *(match)*.

Similarly $V_j^I(i)$, $V_j^D(i)$ are the scores for state I_j *(insertion)* and D_j *(deletion)* respectively.

Then we can write:

$$V_j^{\mathrm{M}}(i) = \log \frac{e_{\mathrm{M}_j}(x_i)}{q_{x_i}} + \max \begin{cases} V_{j-1}^{\mathrm{M}}(i-1) + \log a_{\mathrm{M}_{j-1}\mathrm{M}_j}, \\ V_{j-1}^{\mathrm{I}}(i-1) + \log a_{\mathrm{I}_{j-1}\mathrm{M}_j}, \\ V_{j-1}^{\mathrm{D}}(i-1) + \log a_{\mathrm{D}_{j-1}\mathrm{M}_j}; \end{cases}$$

$$V_j^{\mathrm{I}}(i) = \log \frac{e_{\mathrm{I}_j}(x_i)}{q_{x_i}} + \max \begin{cases} V_j^{\mathrm{M}}(i-1) + \log a_{\mathrm{M}_j\mathrm{I}_j}, \\ V_j^{\mathrm{I}}(i-1) + \log a_{\mathrm{I}_j\mathrm{I}_j}, \\ V_j^{\mathrm{D}}(i-1) + \log a_{\mathrm{D}_j\mathrm{I}_j}; \end{cases}$$

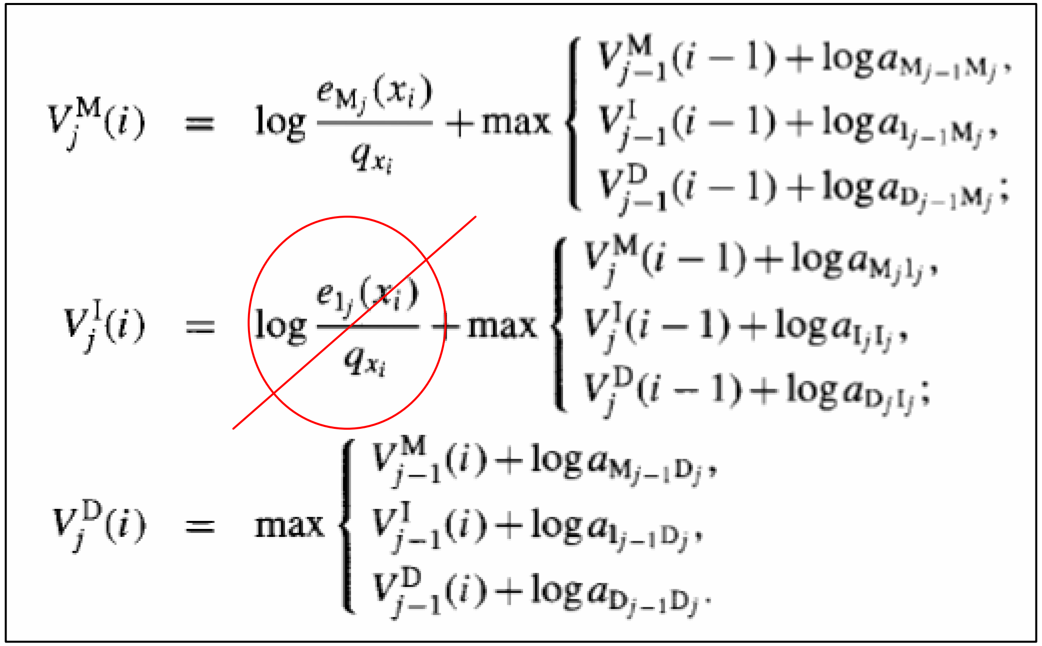
$$V_j^{\mathrm{D}}(i) = \max \begin{cases} V_{j-1}^{\mathrm{M}}(i) + \log a_{\mathrm{M}_{j-1}\mathrm{D}_j}, \\ V_{j-1}^{\mathrm{I}}(i) + \log a_{\mathrm{I}_{j-1}\mathrm{D}_j}, \\ V_{j-1}^{\mathrm{D}}(i) + \log a_{\mathrm{D}_{j-1}\mathrm{D}_j}. \end{cases}$$

Profile HMMs: Viterbi

Let $V_j^M(i)$ be the log-odds score of the best path matching subsequence $x_{1\ldots i}$ to the submodel up to state j, ending with x_i being emitted by state M^j *(match)*.

Similarly $V_j^I(i)$, $V_j^D(i)$ are the scores for state I_j *(insertion)* and D_j *(deletion)* respectively.

Then we can write:



$$V_j^{\mathrm{M}}(i) = \log \frac{e_{\mathrm{M}_j}(x_i)}{q_{x_i}} + \max \begin{cases} V_{j-1}^{\mathrm{M}}(i-1) + \log a_{\mathrm{M}_{j-1}\mathrm{M}_j}, \\ V_{j-1}^{\mathrm{I}}(i-1) + \log a_{\mathrm{I}_{j-1}\mathrm{M}_j}, \\ V_{j-1}^{\mathrm{D}}(i-1) + \log a_{\mathrm{D}_{j-1}\mathrm{M}_j}; \end{cases}$$

$$V_j^{\mathrm{I}}(i) = \cancel{\log \frac{e_{\mathrm{I}_j}(x_i)}{q_{x_i}}} + \max \begin{cases} V_j^{\mathrm{M}}(i-1) + \log a_{\mathrm{M}_j\mathrm{I}_j}, \\ V_j^{\mathrm{I}}(i-1) + \log a_{\mathrm{I}_j\mathrm{I}_j}, \\ V_j^{\mathrm{D}}(i-1) + \log a_{\mathrm{D}_j\mathrm{I}_j}; \end{cases}$$

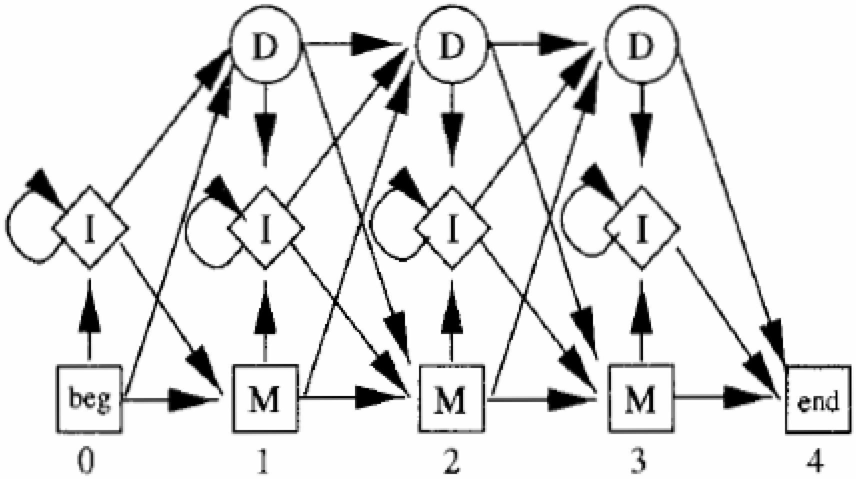
$$V_j^{\mathrm{D}}(i) = \max \begin{cases} V_{j-1}^{\mathrm{M}}(i) + \log a_{\mathrm{M}_{j-1}\mathrm{D}_j}, \\ V_{j-1}^{\mathrm{I}}(i) + \log a_{\mathrm{I}_{j-1}\mathrm{D}_j}, \\ V_{j-1}^{\mathrm{D}}(i) + \log a_{\mathrm{D}_{j-1}\mathrm{D}_j}. \end{cases}$$

Profile HMMs: model construction

(a) Multiple alignment:

	x	x	.	.	.	x
bat	A	G	-	-	-	C
rat	A	-	A	G	-	C
cat	A	G	-	A	A	-
gnat	-	-	A	A	A	C
goat	A	G	-	-	-	C
	1	2	.	.	.	3

(b) Profile-HMM architecture:



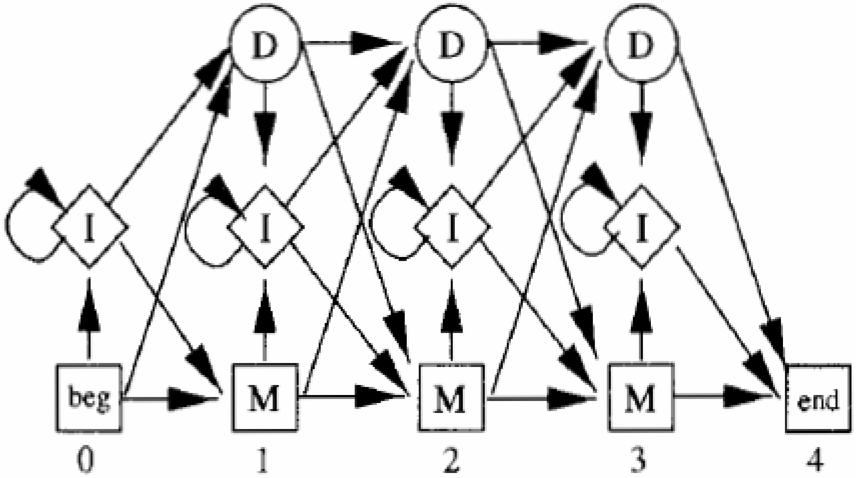
(c) Observed emission/transition counts

		model position			
		0	1	2	3
match emissions	A	-	4	0	0
	C	-	0	0	4
	G	-	0	3	0
	T	-	0	0	0
insert emissions	A	0	0	6	0
	C	0	0	0	0
	G	0	0	1	0
	T	0	0	0	0
state transitions	M-M	4	3	2	4
	M-D	1	1	0	0
	M-I	0	0	1	0
	I-M	0	0	2	0
	I-D	0	0	1	0
	I-I	0	0	4	0
	D-M	-	0	0	1
	D-D	-	1	0	0
	D-I	-	0	2	0

(a) Multiple alignment:

```
      x x . . . x
bat   A G - - - C
rat   A - A G - C
cat   A G - A A -
gnat  - - A A A C
goat  A G - - - C
      1 2 . . . 3
```

(b) Profile-HMM architecture:



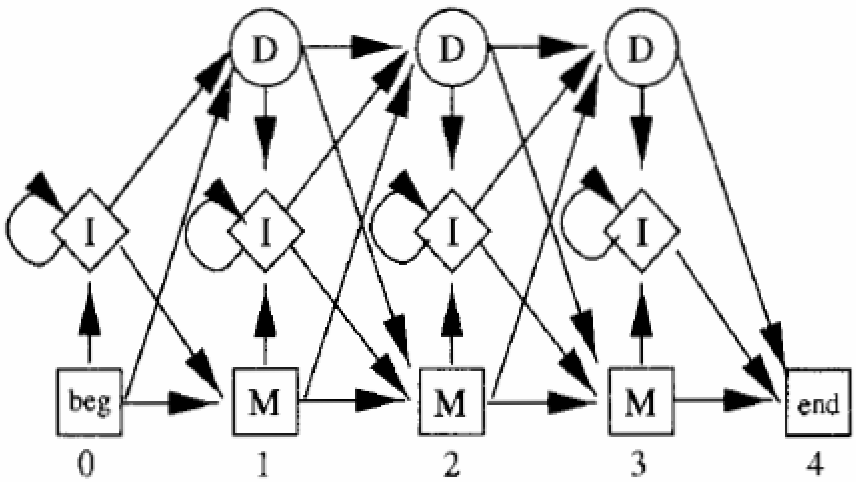
(c) Observed emission/transition counts

		model position 0	1	2	3
match emissions	A	-	4	0	0
	C	-	0	0	4
	G	-	0	3	0
	T	-	0	0	0
insert emissions	A	0	0	6	0
	C	0	0	0	0
	G	0	0	1	0
	T	0	0	0	0
state transitions	M-M	4	3	2	4
	M-D	1	1	0	0
	M-I	0	0	1	0
	I-M	0	0	2	0
	I-D	0	0	1	0
	I-I	0	0	4	0
	D-M	-	0	0	1
	D-D	-	1	0	0
	D-I	-	0	2	0

(a) Multiple alignment:

	x	x	.	.	.	x
bat	A	G	-	-	-	C
rat	A	-	A	G	-	C
cat	A	G	-	A	A	-
gnat	-	-	A	A	A	C
goat	A	G	-	-	-	C
	1	2	.	.	.	3

(b) Profile-HMM architecture:



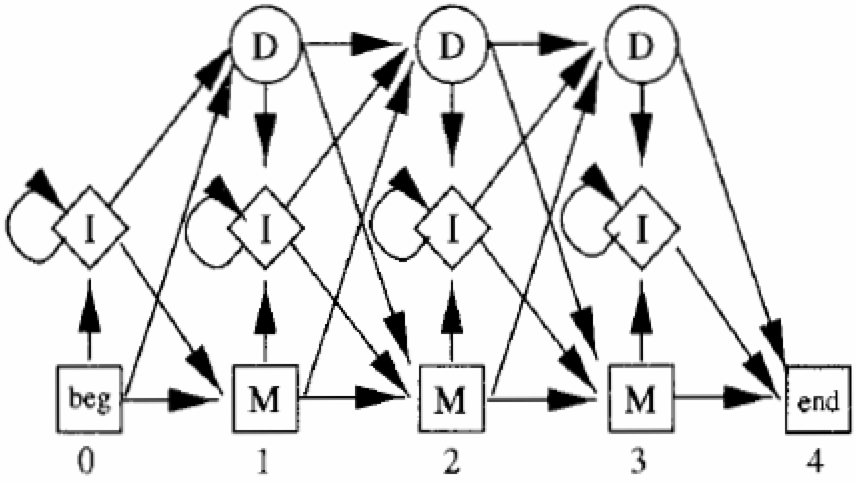
(c) Observed emission/transition counts

		model position			
		0	1	2	3
match emissions	A	-	4	0	0
	C	-	0	0	4
	G	-	0	3	0
	T	-	0	0	0
insert emissions	A	0	0	6	0
	C	0	0	0	0
	G	0	0	1	0
	T	0	0	0	0
state transitions	M-M	4	3	2	4
	M-D	1	1	0	0
	M-I	0	0	1	0
	I-M	0	0	2	0
	I-D	0	0	1	0
	I-I	0	0	4	0
	D-M	-	0	0	1
	D-D	-	1	0	0
	D-I	-	0	2	0

(a) Multiple alignment:

```
      x  x  .  .  .  x
bat   A  G  -  -  -  C
rat   A  -  A  G  -  C
cat   A  G  -  A  A  -
gnat  -  -  A  A  A  C
goat  A  G  -  -  -  C
      1  2  .  .  .  3
```

(b) Profile-HMM architecture:



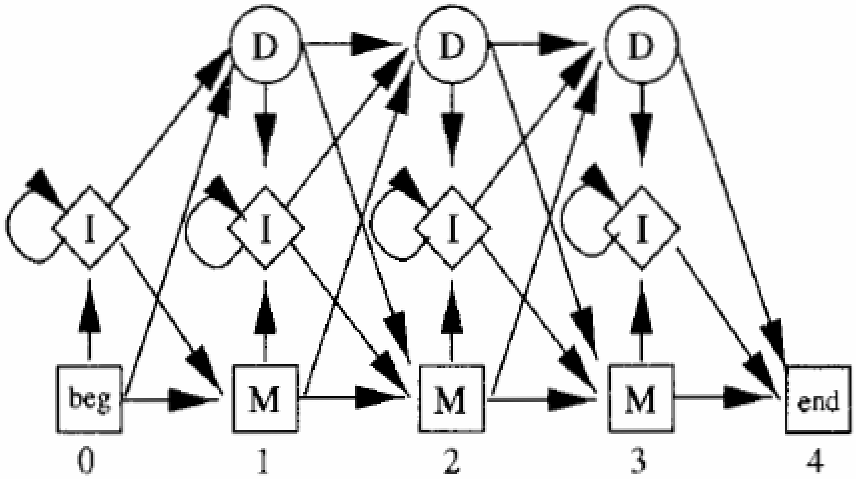
(c) Observed emission/transition counts

		model position 0	1	2	3
match emissions	A	-	4	0	0
	C	-	0	0	4
	G	-	0	3	0
	T	-	0	0	0
insert emissions	A	0	0	6	0
	C	0	0	0	0
	G	0	0	1	0
	T	0	0	0	0
state transitions	M-M	4	3	2	4
	M-D	1	1	0	0
	M-I	0	0	1	0
	I-M	0	0	2	0
	I-D	0	0	1	0
	I-I	0	0	4	0
	D-M	-	0	0	1
	D-D	-	1	0	0
	D-I	-	0	2	0

(a) Multiple alignment:

```
        x x . . . x
bat     A G - - - C
rat     A - A G - C
cat     A G - A A -
gnat    - - A A A C
goat    A G - - - C
        1 2 . . . 3
```

(b) Profile-HMM architecture:

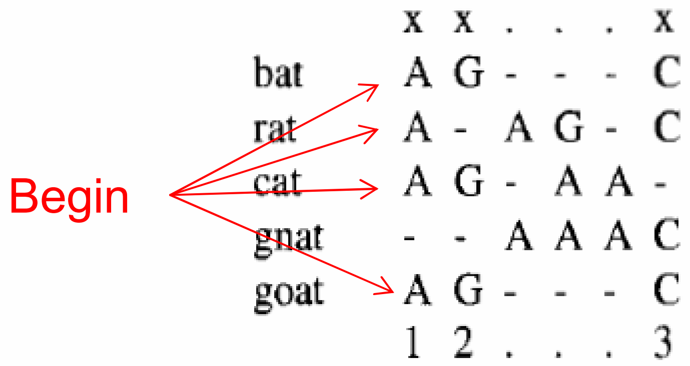


(c) Observed emission/transition counts

model position

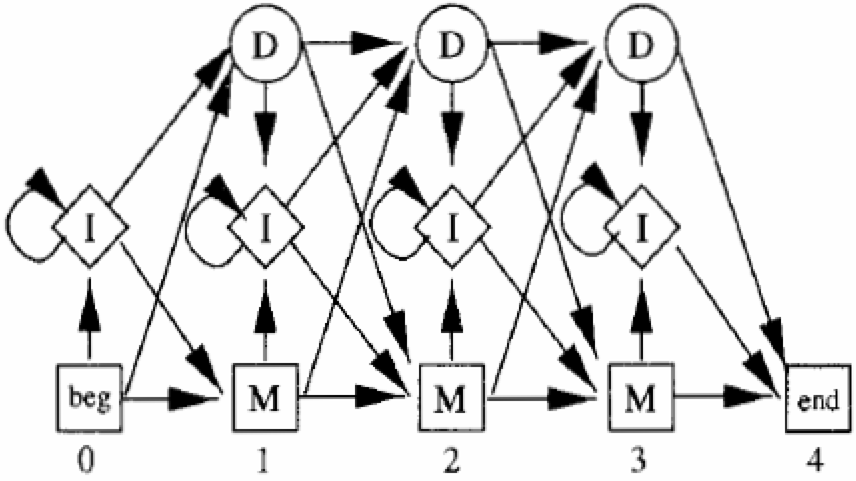
		0	1	2	3
match emissions	A	-	4	0	0
	C	-	0	0	4
	G	-	0	3	0
	T	-	0	0	0
insert emissions	A	0	0	6	0
	C	0	0	0	0
	G	0	0	1	0
	T	0	0	0	0
state transitions	M-M	4	3	2	4
	M-D	1	1	0	0
	M-I	0	0	1	0
	I-M	0	0	2	0
	I-D	0	0	1	0
	I-I	0	0	4	0
	D-M	-	0	0	1
	D-D	-	1	0	0
	D-I	-	0	2	0

(a) Multiple alignment:



```
           x x . . . x
     bat   A G - - - C
     rat   A - A G - C
     cat   A G - A A -
     gnat  - - A A A C
     goat  A G - - - C
           1 2 . . . 3
```

(b) Profile-HMM architecture:



(c) Observed emission/transition counts

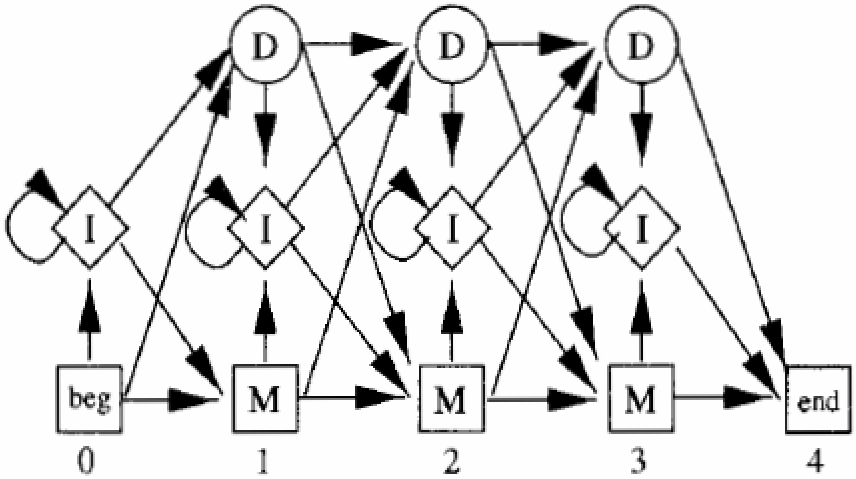
		model position			
		0	1	2	3
match emissions	A	-	4	0	0
	C	-	0	0	4
	G	-	0	3	0
	T	-	0	0	0
insert emissions	A	0	0	6	0
	C	0	0	0	0
	G	0	0	1	0
	T	0	0	0	0
state transitions	M-M	4	3	2	4
	M-D	1	1	0	0
	M-I	0	0	1	0
	I-M	0	0	2	0
	I-D	0	0	1	0
	I-I	0	0	4	0
	D-M	-	0	0	1
	D-D	-	1	0	0
	D-I	-	0	2	0

(a) Multiple alignment:

Begin

	x	x	.	.	.	x
bat	A	G	-	-	-	C
rat	A	-	A	G	-	C
cat	A	G	-	A	A	-
gnat	-	-	A	A	A	C
goat	A	G	-	-	-	C
	1	2	.	.	.	3

(b) Profile-HMM architecture:



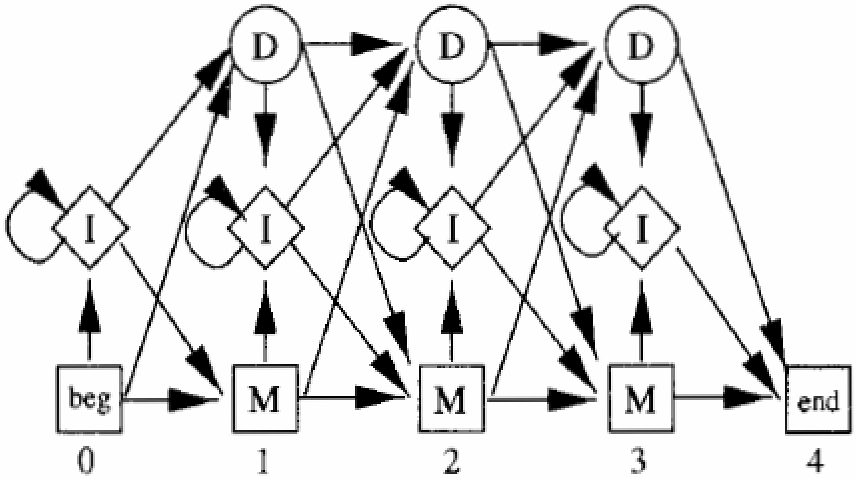
(c) Observed emission/transition counts

		model position			
		0	1	2	3
match emissions	A	-	4	0	0
	C	-	0	0	4
	G	-	0	3	0
	T	-	0	0	0
insert emissions	A	0	0	6	0
	C	0	0	0	0
	G	0	0	1	0
	T	0	0	0	0
state transitions	M-M	4	3	2	4
	M-D	1	1	0	0
	M-I	0	0	1	0
	I-M	0	0	2	0
	I-D	0	0	1	0
	I-I	0	0	4	0
	D-M	-	0	0	1
	D-D	-	1	0	0
	D-I	-	0	2	0

(a) Multiple alignment:

```
        x  x  .  .  .  x
bat     A  G  -  -  -  C
rat     A  -  A  G  -  C
cat     A  G  -  A  A  -
gnat    -  -  A  A  A  C
goat    A  G  -  -  -  C
        1  2  .  .  .  3
```

(b) Profile-HMM architecture:



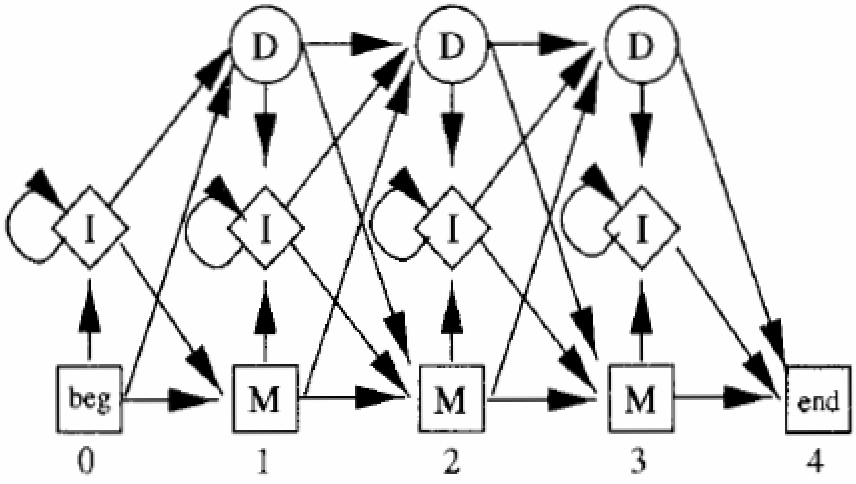
(c) Observed emission/transition counts

		model position			
		0	1	2	3
match emissions	A	-	4	0	0
	C	-	0	0	4
	G	-	0	3	0
	T	-	0	0	0
insert emissions	A	0	0	6	0
	C	0	0	0	0
	G	0	0	1	0
	T	0	0	0	0
state transitions	M-M	4	3	2	4
	M-D	1	1	0	0
	M-I	0	0	1	0
	I-M	0	0	2	0
	I-D	0	0	1	0
	I-I	0	0	4	0
	D-M	-	0	0	1
	D-D	-	1	0	0
	D-I	-	0	2	0

(a) Multiple alignment:

```
        x x . . . x
bat     A G - - - C
rat     A - A G - C
cat     A G - A A -
gnat    - - A A A C
goat    A G - - - C
        1 2 . . . 3
```

(b) Profile-HMM architecture:



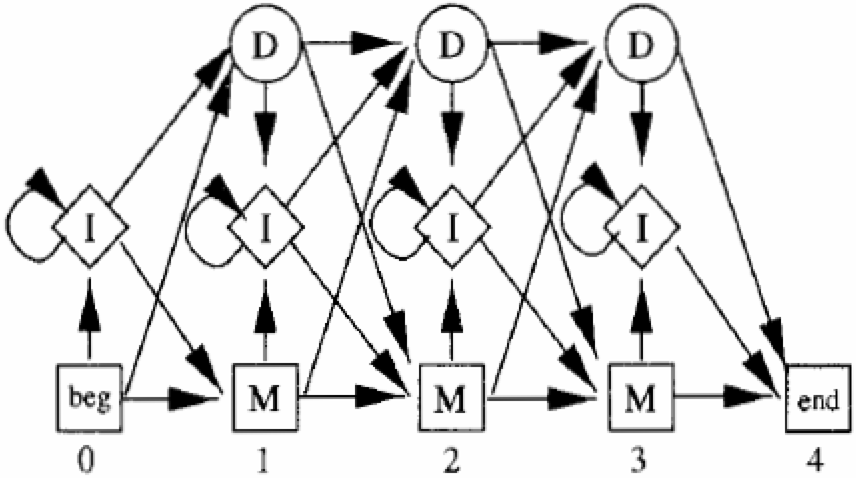
(c) Observed emission/transition counts

		model position 0	1	2	3
match emissions	A	-	4	0	0
	C	-	0	0	4
	G	-	0	3	0
	T	-	0	0	0
insert emissions	A	0	0	6	0
	C	0	0	0	0
	G	0	0	1	0
	T	0	0	0	0
state transitions	M-M	4	3	2	4
	M-D	1	1	0	0
	M-I	0	0	1	0
	I-M	0	0	2	0
	I-D	0	0	1	0
	I-I	0	0	4	0
	D-M	-	0	0	1
	D-D	-	1	0	0
	D-I	-	0	2	0

(a) Multiple alignment:

```
        x x . . . x
bat     A G - - - C
rat     A - A G - C
cat     A G - A A -
gnat    - - A A A C
goat    A G - - - C
        1 2 . . . 3
```

(b) Profile-HMM architecture:



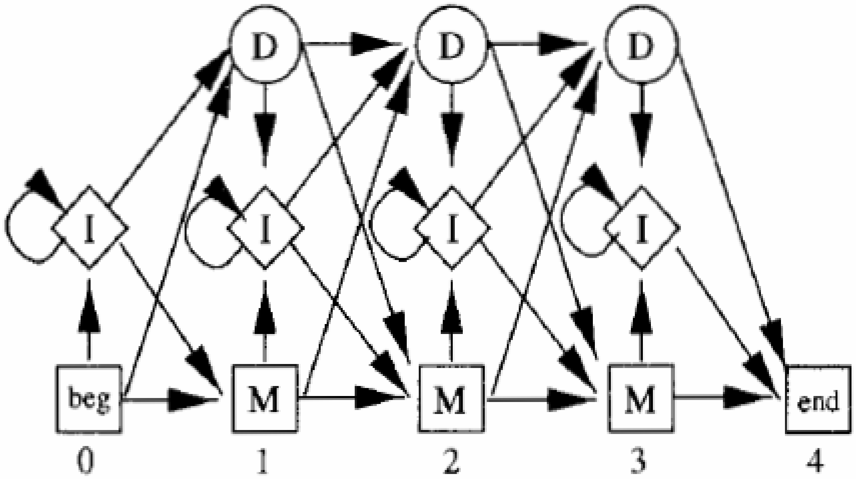
(c) Observed emission/transition counts

		model position 0	1	2	3
match emissions	A	-	4	0	0
	C	-	0	0	4
	G	-	0	3	0
	T	-	0	0	0
insert emissions	A	0	0	6	0
	C	0	0	0	0
	G	0	0	1	0
	T	0	0	0	0
state transitions	M-M	4	3	2	4
	M-D	1	1	0	0
	M-I	0	0	1	0
	I-M	0	0	2	0
	I-D	0	0	1	0
	I-I	0	0	4	0
	D-M	-	0	0	1
	D-D	-	1	0	0
	D-I	-	0	2	0

(a) Multiple alignment:

```
      x x . . . x
bat   A G - - - C
rat   A - A G - C
cat   A G - A A -
gnat  - - A A A C
goat  A G - - - C
      1 2 . . . 3
```

(b) Profile-HMM architecture:



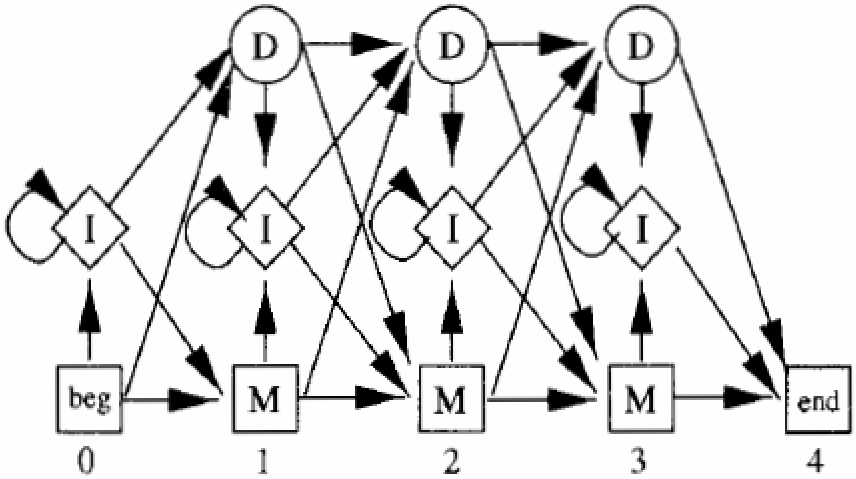
(c) Observed emission/transition counts

		model position 0	1	2	3
match emissions	A	-	4	0	0
	C	-	0	0	4
	G	-	0	3	0
	T	-	0	0	0
insert emissions	A	0	0	6	0
	C	0	0	0	0
	G	0	0	1	0
	T	0	0	0	0
state transitions	M-M	4	3	2	4
	M-D	1	1	0	0
	M-I	0	0	1	0
	I-M	0	0	2	0
	I-D	0	0	1	0
	I-I	0	0	4	0
	D-M	-	0	0	1
	D-D	-	1	0	0
	D-I	-	0	2	0

(a) Multiple alignment:

```
        x x . . . x
bat     A G - - - C
rat     A - A G - C
cat     A G - A A -
gnat    - - A A A C
goat    A G - - - C
        1 2 . . . 3
```

(b) Profile-HMM architecture:



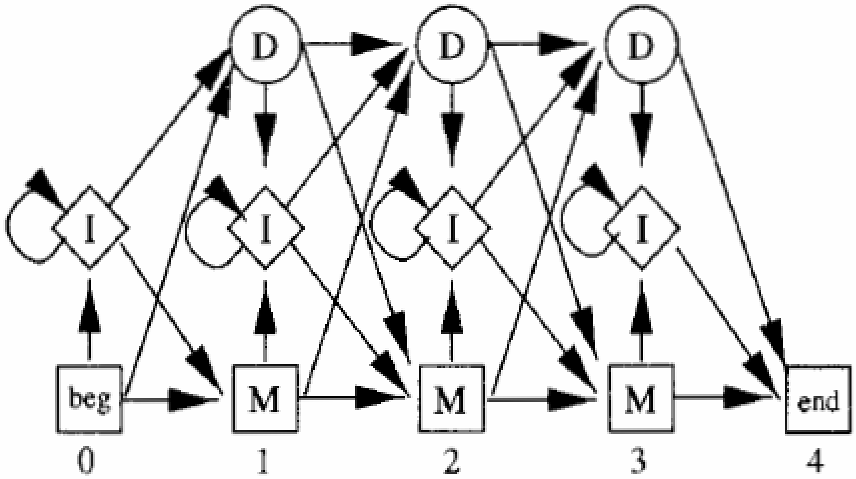
(c) Observed emission/transition counts

		model position			
		0	1	2	3
match emissions	A	-	4	0	0
	C	-	0	0	4
	G	-	0	3	0
	T	-	0	0	0
insert emissions	A	0	0	6	0
	C	0	0	0	0
	G	0	0	1	0
	T	0	0	0	0
state transitions	M-M	4	3	2	4
	M-D	1	1	0	0
	M-I	0	0	1	0
	I-M	0	0	2	0
	I-D	0	0	1	0
	I-I	0	0	4	0
	D-M	-	0	0	1
	D-D	-	1	0	0
	D-I	-	0	2	0

(a) Multiple alignment:

```
       x x . . . x
bat    A G - - - C
rat    A - A G - C
cat    A G - A A -
gnat   - - A A A C
goat   A G - - - C
       1 2 . . . 3
```

(b) Profile-HMM architecture:



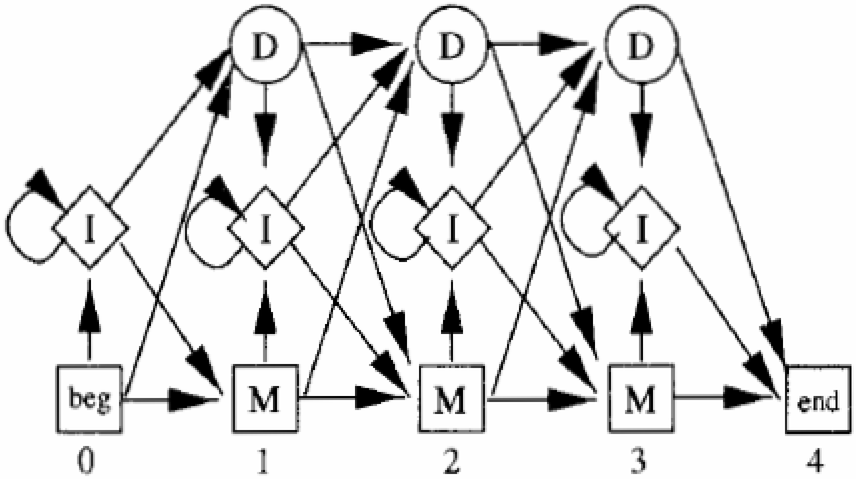
(c) Observed emission/transition counts

		model position 0	1	2	3
match emissions	A	-	4	0	0
	C	-	0	0	4
	G	-	0	3	0
	T	-	0	0	0
insert emissions	A	0	0	6	0
	C	0	0	0	0
	G	0	0	1	0
	T	0	0	0	0
state transitions	M-M	4	3	2	4
	M-D	1	1	0	0
	M-I	0	0	1	0
	I-M	0	0	2	0
	I-D	0	0	1	0
	I-I	0	0	4	0
	D-M	-	0	0	1
	D-D	-	1	0	0
	D-I	-	0	2	0

(a) Multiple alignment:

```
      x  x  .  .  .  x
bat   A  G  -  -  -  C
rat   A  -  A  G  -  C
cat   A  G  -  A  A  -
gnat  -  -  A  A  A  C
goat  A  G  -  -  -  C
      1  2  .  .  .  3
```

(b) Profile-HMM architecture:



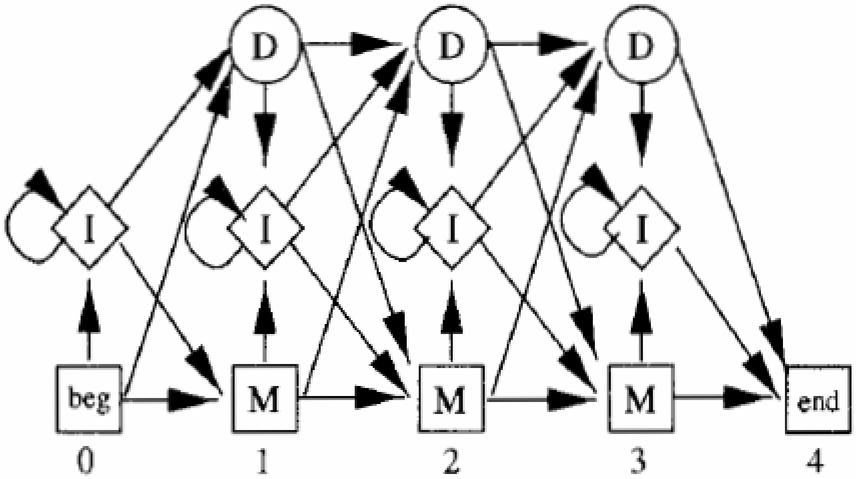
(c) Observed emission/transition counts

		model position 0	1	2	3
match emissions	A	-	4	0	0
	C	-	0	0	4
	G	-	0	3	0
	T	-	0	0	0
insert emissions	A	0	0	6	0
	C	0	0	0	0
	G	0	0	1	0
	T	0	0	0	0
state transitions	M-M	4	3	2	4
	M-D	1	1	0	0
	M-I	0	0	1	0
	I-M	0	0	2	0
	I-D	0	0	1	0
	I-I	0	0	4	0
	D-M	-	0	0	1
	D-D	-	1	0	0
	D-I	-	0	2	0

(a) Multiple alignment:

```
      x x . . . x
bat   A G - - - C
rat   A - A G - C
cat   A G - A A -
gnat  - - A A A C
goat  A G - - - C
      1 2 . . . 3
```

(b) Profile-HMM architecture:



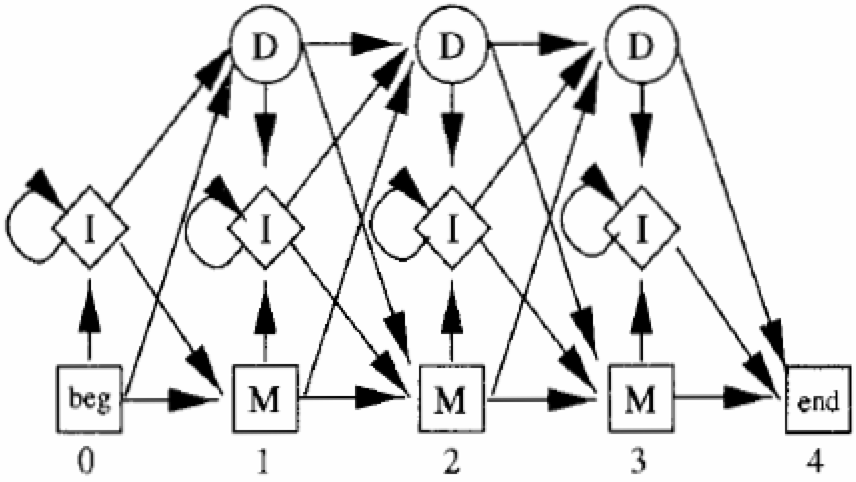
(c) Observed emission/transition counts

		model position 0	1	2	3
match emissions	A	-	4	0	0
	C	-	0	0	4
	G	-	0	3	0
	T	-	0	0	0
insert emissions	A	0	0	6	0
	C	0	0	0	0
	G	0	0	1	0
	T	0	0	0	0
state transitions	M-M	4	3	2	4
	M-D	1	1	0	0
	M-I	0	0	1	0
	I-M	0	0	2	0
	I-D	0	0	1	0
	I-I	0	0	4	0
	D-M	-	0	0	1
	D-D	-	1	0	0
	D-I	-	0	2	0

(a) Multiple alignment:

	x	x	.	.	.	x
bat	A	G	-	-	-	C
rat	A	-	A	G	-	C
cat	A	G	-	A	A	-
gnat	-	-	A	A	A	C
goat	A	G	-	-	-	C
	1	2	.	.	.	3

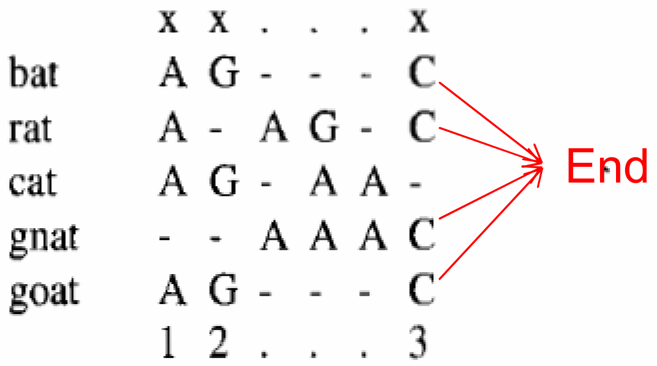
(b) Profile-HMM architecture:



(c) Observed emission/transition counts

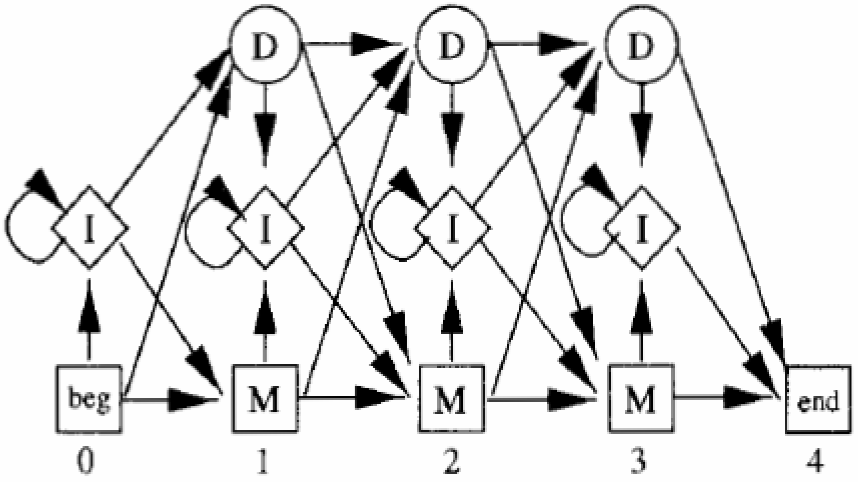
		model position			
		0	1	2	3
match emissions	A	-	4	0	0
	C	-	0	0	4
	G	-	0	3	0
	T	-	0	0	0
insert emissions	A	0	0	6	0
	C	0	0	0	0
	G	0	0	1	0
	T	0	0	0	0
state transitions	M-M	4	3	2	4
	M-D	1	1	0	0
	M-I	0	0	1	0
	I-M	0	0	2	0
	I-D	0	0	1	0
	I-I	0	0	4	0
	D-M	-	0	0	1
	D-D	-	1	0	0
	D-I	-	0	2	0

(a) Multiple alignment:



```
       x x . . . x
bat    A G - - - C
rat    A - A G - C
cat    A G - A A -
gnat   - - A A A C
goat   A G - - - C
       1 2 . . . 3
```

(b) Profile-HMM architecture:



(c) Observed emission/transition counts

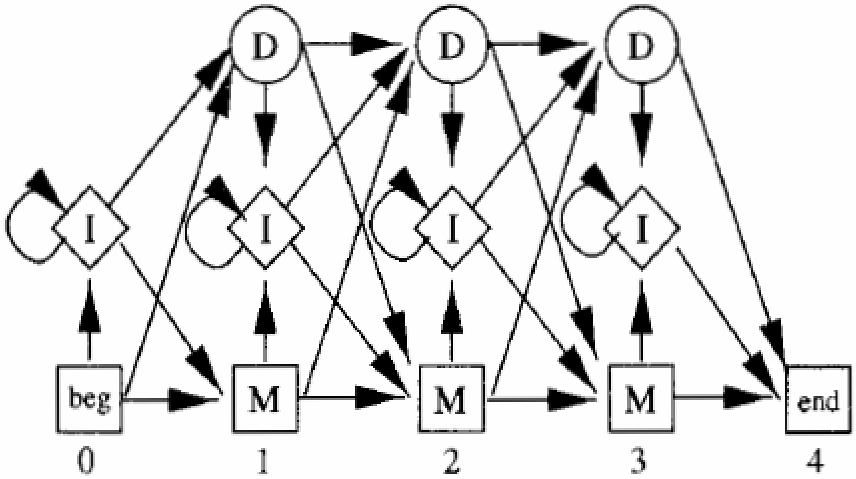
		model position 0	1	2	3
match emissions	A	-	4	0	0
	C	-	0	0	4
	G	-	0	3	0
	T	-	0	0	0
insert emissions	A	0	0	6	0
	C	0	0	0	0
	G	0	0	1	0
	T	0	0	0	0
state transitions	M-M	4	3	2	4
	M-D	1	1	0	0
	M-I	0	0	1	0
	I-M	0	0	2	0
	I-D	0	0	1	0
	I-I	0	0	4	0
	D-M	-	0	0	1
	D-D	-	1	0	0
	D-I	-	0	2	0

(a) Multiple alignment:

```
       x x . . . x
bat    A G - - - C
rat    A - A G - C
cat    A G - A A -
gnat   - - A A A C
goat   A G - - - C
       1 2 . . . 3
```

End

(b) Profile-HMM architecture:



(c) Observed emission/transition counts

		model position 0	1	2	3
match emissions	A	-	4	0	0
	C	-	0	0	4
	G	-	0	3	0
	T	-	0	0	0
insert emissions	A	0	0	6	0
	C	0	0	0	0
	G	0	0	1	0
	T	0	0	0	0
state transitions	M-M	4	3	2	4
	M-D	1	1	0	0
	M-I	0	0	1	0
	I-M	0	0	2	0
	I-D	0	0	1	0
	I-I	0	0	4	0
	D-M	-	0	0	1
	D-D	-	1	0	0
	D-I	-	0	2	0

HMMER

hmmalign - align sequences to a profile HMM

SOURCE: http://hmmer.janelia.org/

HMMER

hmmalign - align sequences to a profile HMM

hmmbuild - construct profile HMM(s) from multiple sequence alignment(s)

SOURCE: http://hmmer.janelia.org/

HMMER

hmmalign - align sequences to a profile HMM

hmmbuild - construct profile HMM(s) from multiple sequence alignment(s)

hmmconvert - convert profile file to a HMMER format

SOURCE: http://hmmer.janelia.org/

HMMER

hmmalign - align sequences to a profile HMM

hmmbuild - construct profile HMM(s) from multiple sequence alignment(s)

hmmconvert - convert profile file to a HMMER format

hmmemit - sample sequences from a profile HMM

SOURCE: http://hmmer.janelia.org/

HMMER

hmmalign - align sequences to a profile HMM

hmmbuild - construct profile HMM(s) from multiple sequence alignment(s)

hmmconvert - convert profile file to a HMMER format

hmmemit - sample sequences from a profile HMM

hmmfetch - retrieve profile HMM(s) from a file

SOURCE: http://hmmer.janelia.org/

HMMER

hmmalign - align sequences to a profile HMM

hmmbuild - construct profile HMM(s) from multiple sequence alignment(s)

hmmconvert - convert profile file to a HMMER format

hmmemit - sample sequences from a profile HMM

hmmfetch - retrieve profile HMM(s) from a file

hmmpress - prepare an HMM database for hmmscan

SOURCE: http://hmmer.janelia.org/

HMMER

hmmalign - align sequences to a profile HMM

hmmbuild - construct profile HMM(s) from multiple sequence alignment(s)

hmmconvert - convert profile file to a HMMER format

hmmemit - sample sequences from a profile HMM

hmmfetch - retrieve profile HMM(s) from a file

hmmpress - prepare an HMM database for hmmscan

hmmscan - search sequence(s) against a profile database

SOURCE: http://hmmer.janelia.org/

HMMER

hmmalign - align sequences to a profile HMM

hmmbuild - construct profile HMM(s) from multiple sequence alignment(s)

hmmconvert - convert profile file to a HMMER format

hmmemit - sample sequences from a profile HMM

hmmfetch - retrieve profile HMM(s) from a file

hmmpress - prepare an HMM database for hmmscan

hmmscan - search sequence(s) against a profile database

hmmsearch - search profile(s) against a sequence database

SOURCE: http://hmmer.janelia.org/

HMMER

hmmalign - align sequences to a profile HMM

hmmbuild - construct profile HMM(s) from multiple sequence alignment(s)

hmmconvert - convert profile file to a HMMER format

hmmemit - sample sequences from a profile HMM

hmmfetch - retrieve profile HMM(s) from a file

hmmpress - prepare an HMM database for hmmscan

hmmscan - search sequence(s) against a profile database

hmmsearch - search profile(s) against a sequence database

hmmsim - collect score distributions on random sequences

SOURCE: http://hmmer.janelia.org/

HMMER

hmmalign - align sequences to a profile HMM

hmmbuild - construct profile HMM(s) from multiple sequence alignment(s)

hmmconvert - convert profile file to a HMMER format

hmmemit - sample sequences from a profile HMM

hmmfetch - retrieve profile HMM(s) from a file

hmmpress - prepare an HMM database for hmmscan

hmmscan - search sequence(s) against a profile database

hmmsearch - search profile(s) against a sequence database

hmmsim - collect score distributions on random sequences

hmmstat - display summary statistics for a profile file

SOURCE: http://hmmer.janelia.org/

HMMER

hmmalign - align sequences to a profile HMM

hmmbuild - construct profile HMM(s) from multiple sequence alignment(s)

hmmconvert - convert profile file to a HMMER format

hmmemit - sample sequences from a profile HMM

hmmfetch - retrieve profile HMM(s) from a file

hmmpress - prepare an HMM database for hmmscan

hmmscan - search sequence(s) against a profile database

hmmsearch - search profile(s) against a sequence database

hmmsim - collect score distributions on random sequences

hmmstat - display summary statistics for a profile file

jackhmmer - iteratively search sequence(s) against a protein database

SOURCE: http://hmmer.janelia.org/

HMMER

hmmalign - align sequences to a profile HMM

hmmbuild - construct profile HMM(s) from multiple sequence alignment(s)

hmmconvert - convert profile file to a HMMER format

hmmemit - sample sequences from a profile HMM

hmmfetch - retrieve profile HMM(s) from a file

hmmpress - prepare an HMM database for hmmscan

hmmscan - search sequence(s) against a profile database

hmmsearch - search profile(s) against a sequence database

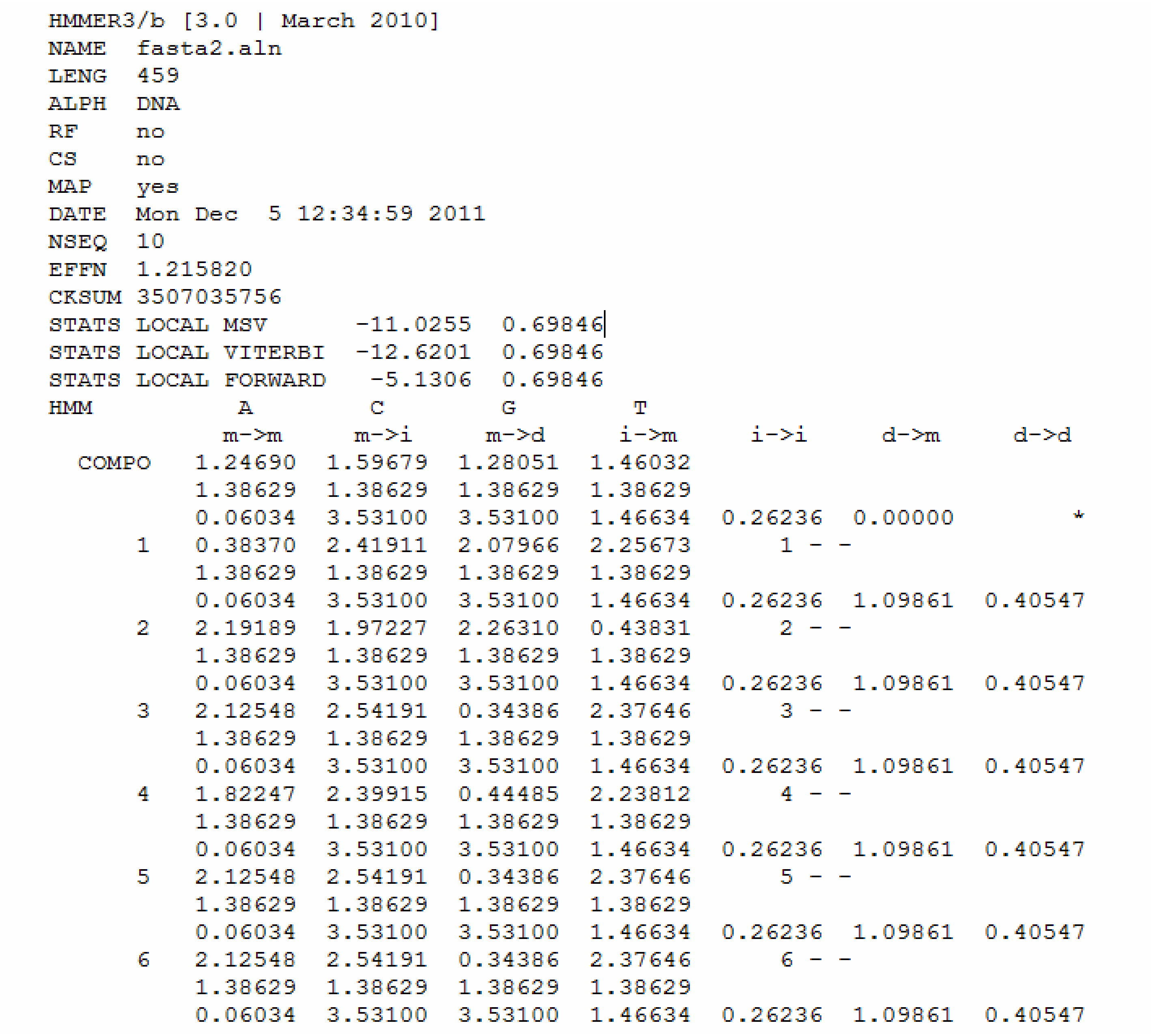
hmmsim - collect score distributions on random sequences

hmmstat - display summary statistics for a profile file

jackhmmer - iteratively search sequence(s) against a protein database

phmmer - search protein sequence(s) against a protein sequence database

SOURCE: http://hmmer.janelia.org/

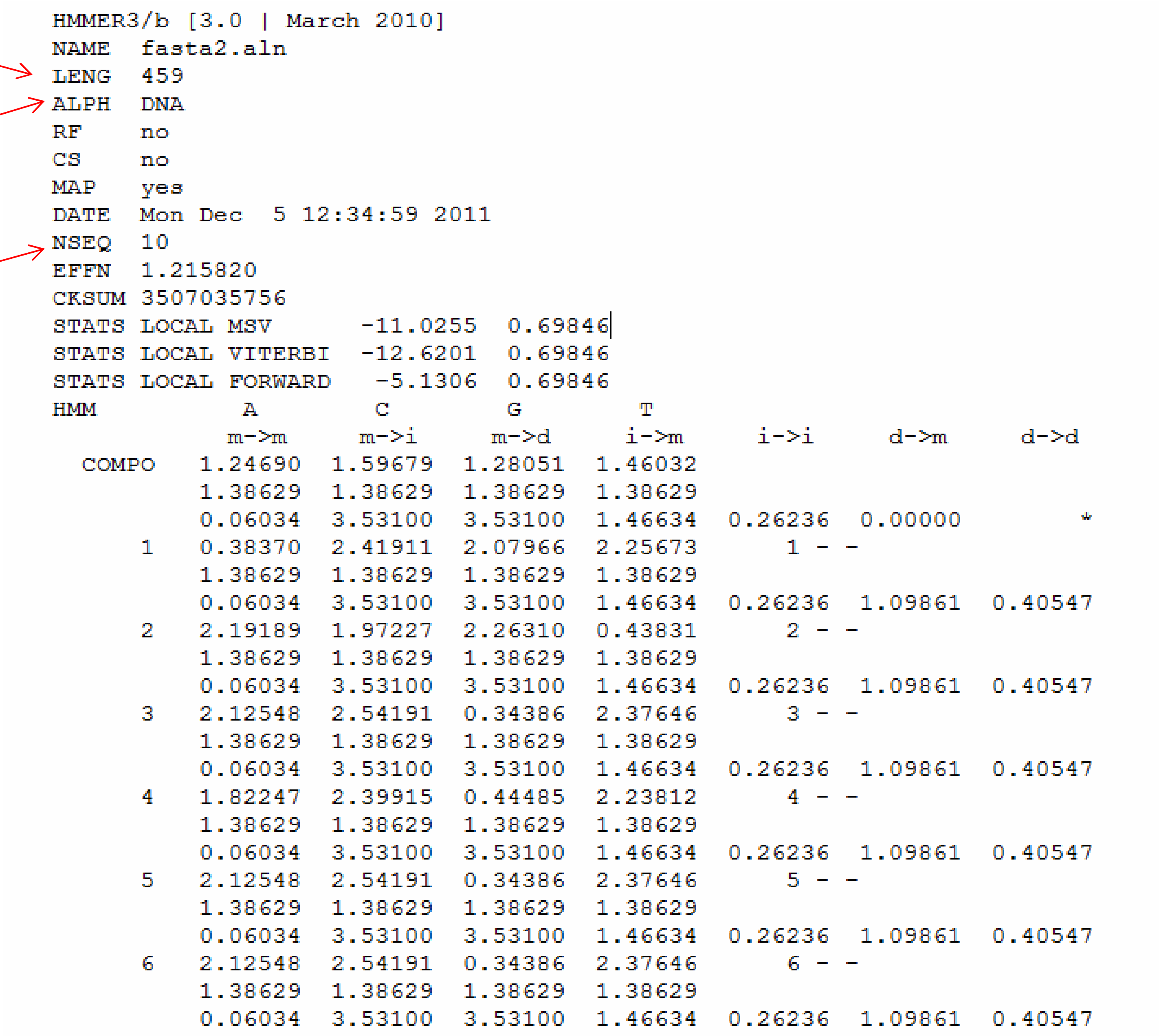


```
HMMER3/b [3.0 | March 2010]
NAME  fasta2.aln
LENG  459
ALPH  DNA
RF    no
CS    no
MAP   yes
DATE  Mon Dec  5 12:34:59 2011
NSEQ  10
EFFN  1.215820
CKSUM 3507035756
STATS LOCAL MSV      -11.0255  0.69846
STATS LOCAL VITERBI  -12.6201  0.69846
STATS LOCAL FORWARD   -5.1306  0.69846
HMM          A        C        G        T
            m->m     m->i     m->d     i->m     i->i     d->m     d->d
  COMPO   1.24690  1.59679  1.28051  1.46032
          1.38629  1.38629  1.38629  1.38629
          0.06034  3.53100  3.53100  1.46634  0.26236  0.00000        *
      1   0.38370  2.41911  2.07966  2.25673      1 - -
          1.38629  1.38629  1.38629  1.38629
          0.06034  3.53100  3.53100  1.46634  0.26236  1.09861  0.40547
      2   2.19189  1.97227  2.26310  0.43831      2 - -
          1.38629  1.38629  1.38629  1.38629
          0.06034  3.53100  3.53100  1.46634  0.26236  1.09861  0.40547
      3   2.12548  2.54191  0.34386  2.37646      3 - -
          1.38629  1.38629  1.38629  1.38629
          0.06034  3.53100  3.53100  1.46634  0.26236  1.09861  0.40547
      4   1.82247  2.39915  0.44485  2.23812      4 - -
          1.38629  1.38629  1.38629  1.38629
          0.06034  3.53100  3.53100  1.46634  0.26236  1.09861  0.40547
      5   2.12548  2.54191  0.34386  2.37646      5 - -
          1.38629  1.38629  1.38629  1.38629
          0.06034  3.53100  3.53100  1.46634  0.26236  1.09861  0.40547
      6   2.12548  2.54191  0.34386  2.37646      6 - -
          1.38629  1.38629  1.38629  1.38629
          0.06034  3.53100  3.53100  1.46634  0.26236  1.09861  0.40547
```

the number of match states in the model.

Symbol alphabet type

the number of sequences that the HMM was trained on



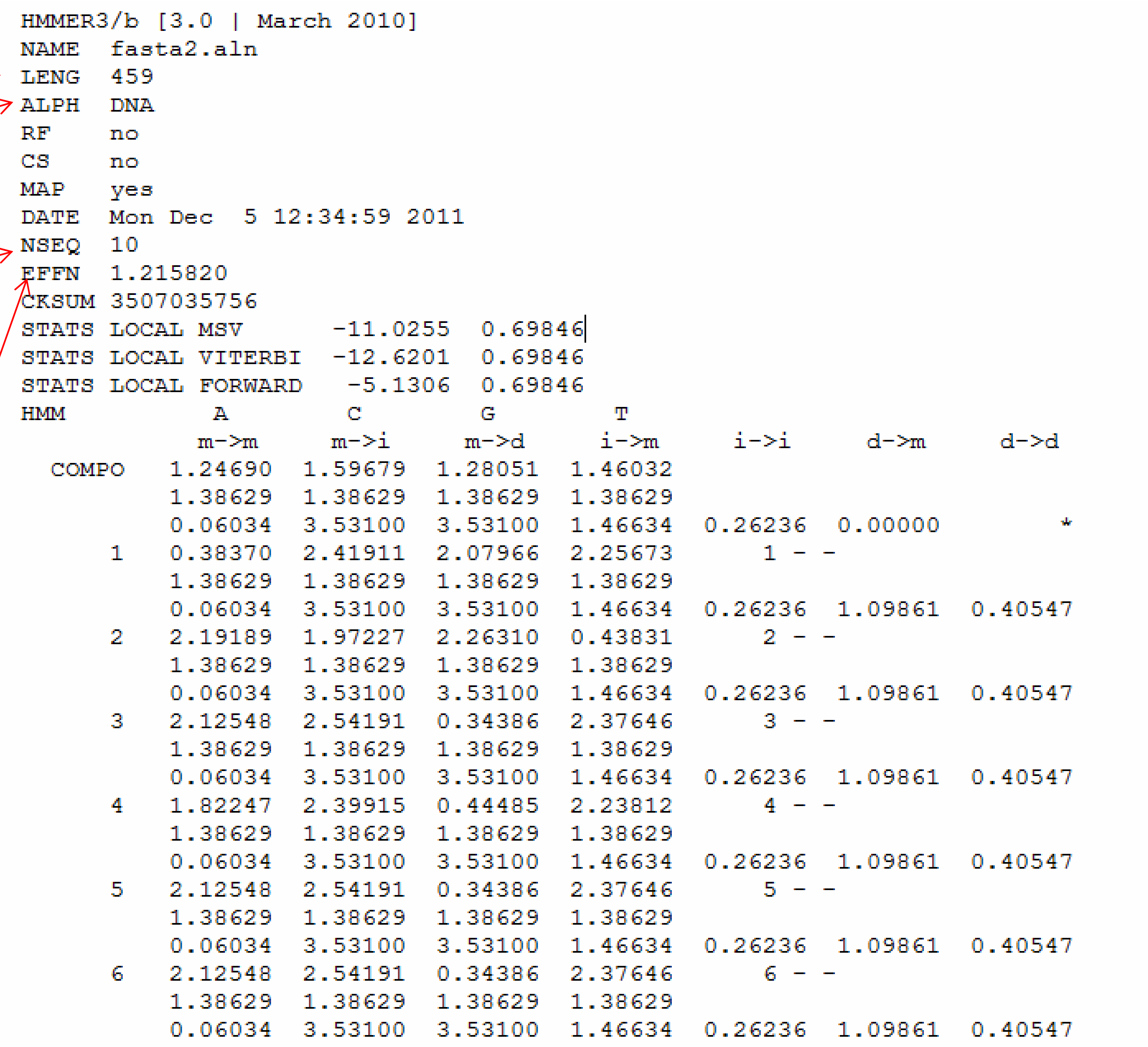
```
HMMER3/b [3.0 | March 2010]
NAME  fasta2.aln
LENG  459
ALPH  DNA
RF    no
CS    no
MAP   yes
DATE  Mon Dec  5 12:34:59 2011
NSEQ  10
EFFN  1.215820
CKSUM 3507035756
STATS LOCAL MSV       -11.0255  0.69846
STATS LOCAL VITERBI   -12.6201  0.69846
STATS LOCAL FORWARD    -5.1306  0.69846
HMM          A        C        G        T
            m->m     m->i     m->d     i->m     i->i     d->m     d->d
  COMPO   1.24690  1.59679  1.28051  1.46032
          1.38629  1.38629  1.38629  1.38629
          0.06034  3.53100  3.53100  1.46634  0.26236  0.00000        *
      1   0.38370  2.41911  2.07966  2.25673      1 - -
          1.38629  1.38629  1.38629  1.38629
          0.06034  3.53100  3.53100  1.46634  0.26236  1.09861  0.40547
      2   2.19189  1.97227  2.26310  0.43831      2 - -
          1.38629  1.38629  1.38629  1.38629
          0.06034  3.53100  3.53100  1.46634  0.26236  1.09861  0.40547
      3   2.12548  2.54191  0.34386  2.37646      3 - -
          1.38629  1.38629  1.38629  1.38629
          0.06034  3.53100  3.53100  1.46634  0.26236  1.09861  0.40547
      4   1.82247  2.39915  0.44485  2.23812      4 - -
          1.38629  1.38629  1.38629  1.38629
          0.06034  3.53100  3.53100  1.46634  0.26236  1.09861  0.40547
      5   2.12548  2.54191  0.34386  2.37646      5 - -
          1.38629  1.38629  1.38629  1.38629
          0.06034  3.53100  3.53100  1.46634  0.26236  1.09861  0.40547
      6   2.12548  2.54191  0.34386  2.37646      6 - -
          1.38629  1.38629  1.38629  1.38629
          0.06034  3.53100  3.53100  1.46634  0.26236  1.09861  0.40547
```

the number of match states in the model.

Symbol alphabet type

the number of sequences that the HMM was trained on

the effective total number of sequences determined by hmmbuild during sequence weighting, for combining observed counts with Dirichlet prior information in parameterizing the model



```
HMMER3/b [3.0 | March 2010]
NAME  fasta2.aln
LENG  459
ALPH  DNA
RF    no
CS    no
MAP   yes
DATE  Mon Dec  5 12:34:59 2011
NSEQ  10
EFFN  1.215820
CKSUM 3507035756
STATS LOCAL MSV       -11.0255  0.69846
STATS LOCAL VITERBI   -12.6201  0.69846
STATS LOCAL FORWARD    -5.1306  0.69846
HMM          A        C        G        T
            m->m     m->i     m->d     i->m     i->i     d->m     d->d
  COMPO   1.24690  1.59679  1.28051  1.46032
          1.38629  1.38629  1.38629  1.38629
          0.06034  3.53100  3.53100  1.46634  0.26236  0.00000        *
      1   0.38370  2.41911  2.07966  2.25673      1 - -
          1.38629  1.38629  1.38629  1.38629
          0.06034  3.53100  3.53100  1.46634  0.26236  1.09861  0.40547
      2   2.19189  1.97227  2.26310  0.43831      2 - -
          1.38629  1.38629  1.38629  1.38629
          0.06034  3.53100  3.53100  1.46634  0.26236  1.09861  0.40547
      3   2.12548  2.54191  0.34386  2.37646      3 - -
          1.38629  1.38629  1.38629  1.38629
          0.06034  3.53100  3.53100  1.46634  0.26236  1.09861  0.40547
      4   1.82247  2.39915  0.44485  2.23812      4 - -
          1.38629  1.38629  1.38629  1.38629
          0.06034  3.53100  3.53100  1.46634  0.26236  1.09861  0.40547
      5   2.12548  2.54191  0.34386  2.37646      5 - -
          1.38629  1.38629  1.38629  1.38629
          0.06034  3.53100  3.53100  1.46634  0.26236  1.09861  0.40547
      6   2.12548  2.54191  0.34386  2.37646      6 - -
          1.38629  1.38629  1.38629  1.38629
          0.06034  3.53100  3.53100  1.46634  0.26236  1.09861  0.40547
```

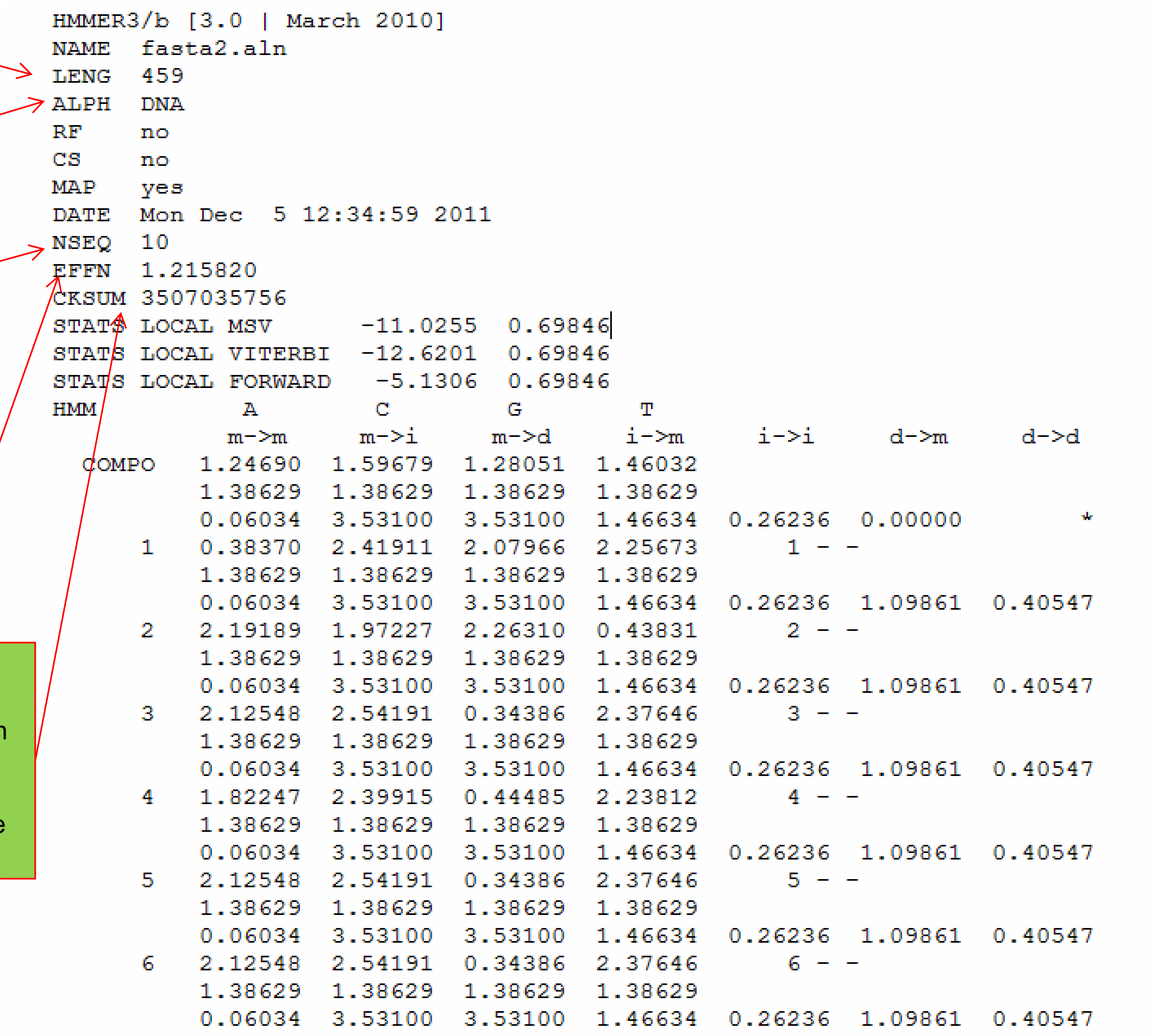
the number of match states in the model.

Symbol alphabet type

the number of sequences that the HMM was trained on

the effective total number of sequences determined by hmmbuild during sequence weighting, for combining observed counts with Dirichlet prior information in parameterizing the model

This number is calculated from the training sequence data, and used in conjunction with the alignment map information to verify that a given alignment is indeed the alignment that the map is for



```
HMMER3/b [3.0 | March 2010]
NAME  fasta2.aln
LENG  459
ALPH  DNA
RF    no
CS    no
MAP   yes
DATE  Mon Dec  5 12:34:59 2011
NSEQ  10
EFFN  1.215820
CKSUM 3507035756
STATS LOCAL MSV       -11.0255  0.69846
STATS LOCAL VITERBI   -12.6201  0.69846
STATS LOCAL FORWARD    -5.1306  0.69846
HMM          A        C        G        T
            m->m     m->i     m->d     i->m     i->i     d->m     d->d
  COMPO   1.24690  1.59679  1.28051  1.46032
          1.38629  1.38629  1.38629  1.38629
          0.06034  3.53100  3.53100  1.46634  0.26236  0.00000        *
      1   0.38370  2.41911  2.07966  2.25673      1 - -
          1.38629  1.38629  1.38629  1.38629
          0.06034  3.53100  3.53100  1.46634  0.26236  1.09861  0.40547
      2   2.19189  1.97227  2.26310  0.43831      2 - -
          1.38629  1.38629  1.38629  1.38629
          0.06034  3.53100  3.53100  1.46634  0.26236  1.09861  0.40547
      3   2.12548  2.54191  0.34386  2.37646      3 - -
          1.38629  1.38629  1.38629  1.38629
          0.06034  3.53100  3.53100  1.46634  0.26236  1.09861  0.40547
      4   1.82247  2.39915  0.44485  2.23812      4 - -
          1.38629  1.38629  1.38629  1.38629
          0.06034  3.53100  3.53100  1.46634  0.26236  1.09861  0.40547
      5   2.12548  2.54191  0.34386  2.37646      5 - -
          1.38629  1.38629  1.38629  1.38629
          0.06034  3.53100  3.53100  1.46634  0.26236  1.09861  0.40547
      6   2.12548  2.54191  0.34386  2.37646      6 - -
          1.38629  1.38629  1.38629  1.38629
          0.06034  3.53100  3.53100  1.46634  0.26236  1.09861  0.40547
```

the number of match states in the model.

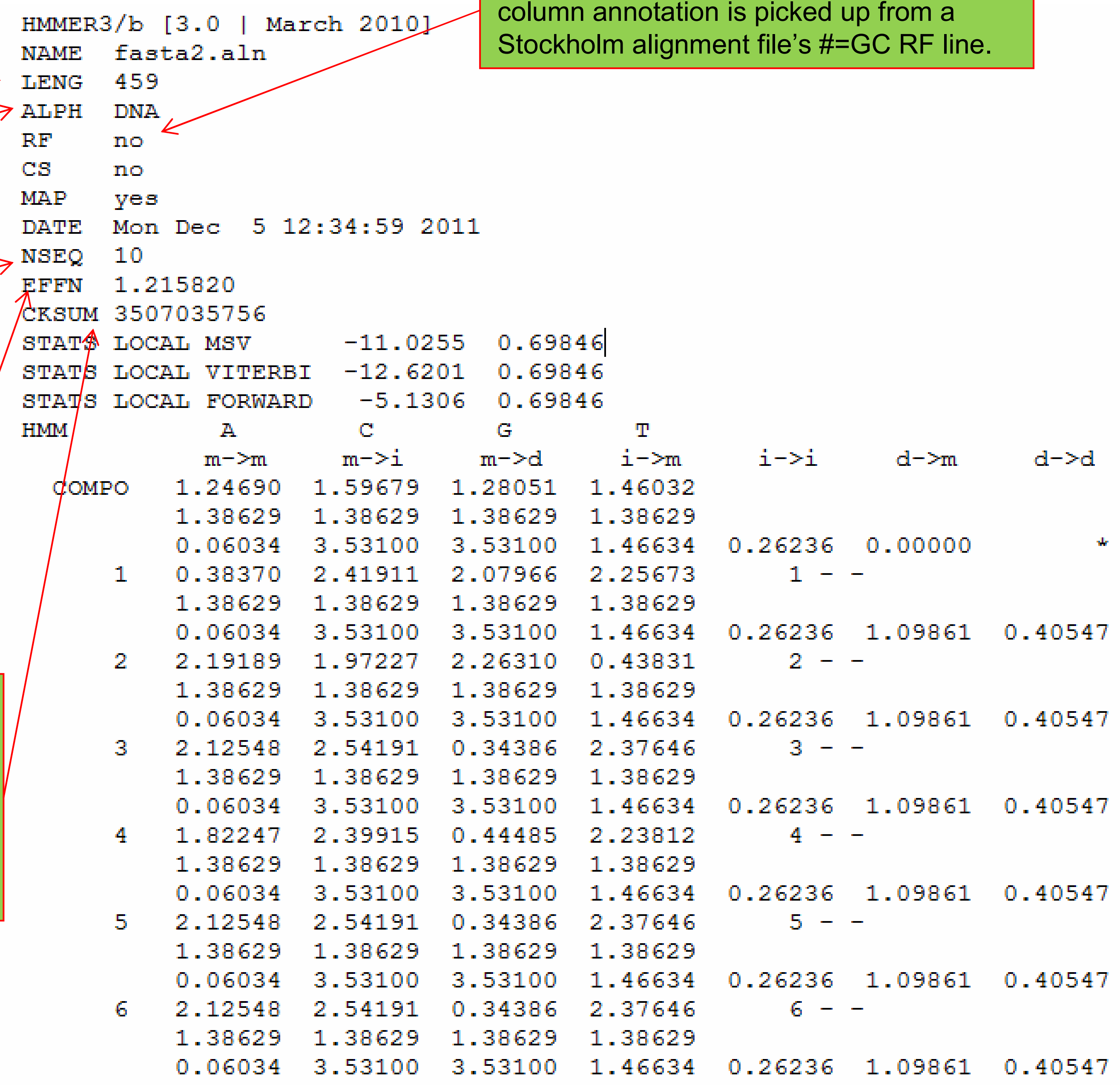
Symbol alphabet type

the number of sequences that the HMM was trained on

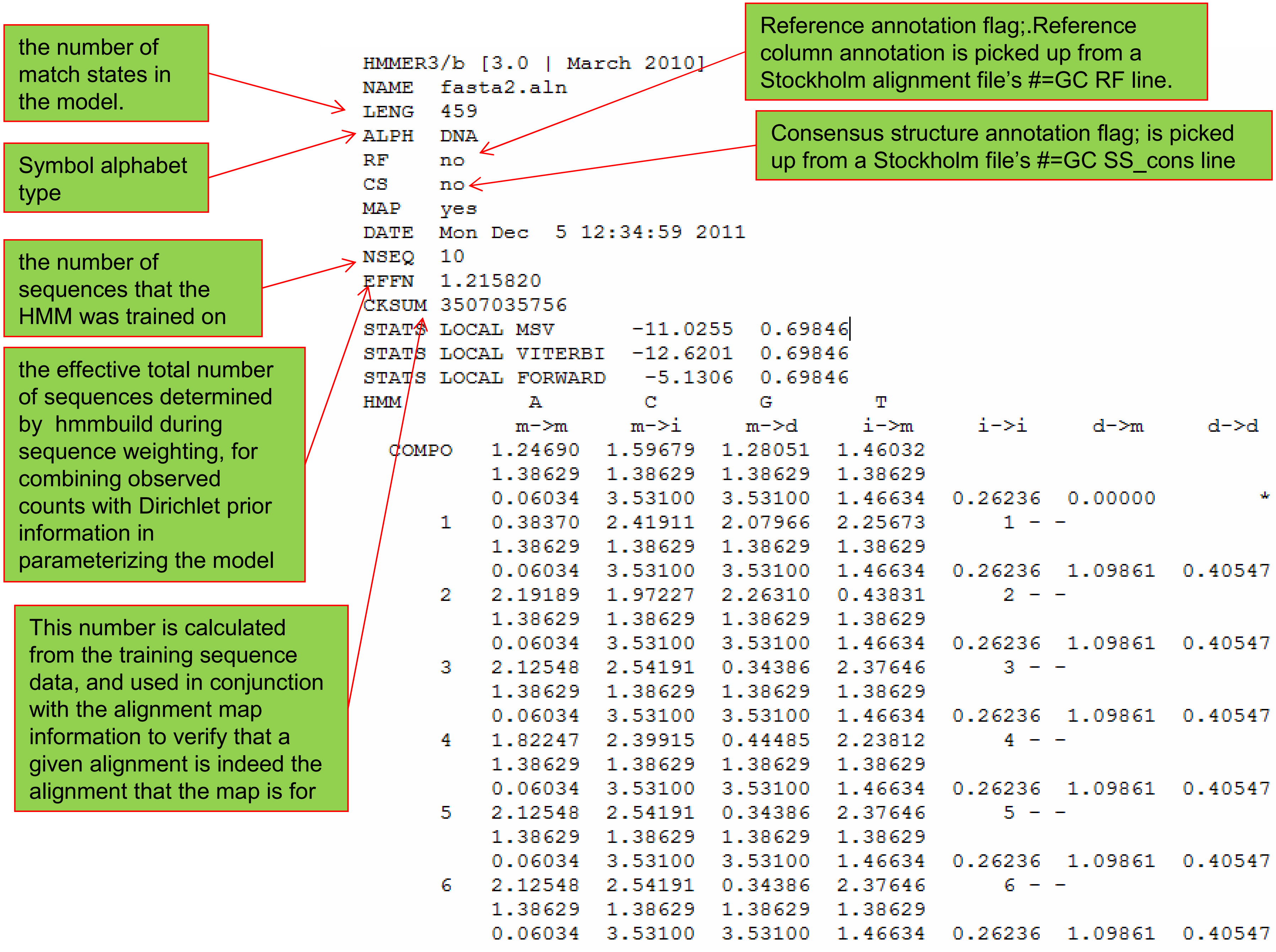
the effective total number of sequences determined by hmmbuild during sequence weighting, for combining observed counts with Dirichlet prior information in parameterizing the model

This number is calculated from the training sequence data, and used in conjunction with the alignment map information to verify that a given alignment is indeed the alignment that the map is for

Reference annotation flag;.Reference column annotation is picked up from a Stockholm alignment file's #=GC RF line.



```
HMMER3/b [3.0 | March 2010]
NAME  fasta2.aln
LENG  459
ALPH  DNA
RF    no
CS    no
MAP   yes
DATE  Mon Dec  5 12:34:59 2011
NSEQ  10
EFFN  1.215820
CKSUM 3507035756
STATS LOCAL MSV       -11.0255  0.69846
STATS LOCAL VITERBI   -12.6201  0.69846
STATS LOCAL FORWARD    -5.1306  0.69846
HMM          A        C        G        T
            m->m     m->i     m->d     i->m     i->i     d->m     d->d
  COMPO   1.24690  1.59679  1.28051  1.46032
          1.38629  1.38629  1.38629  1.38629
          0.06034  3.53100  3.53100  1.46634  0.26236  0.00000        *
      1   0.38370  2.41911  2.07966  2.25673      1 - -
          1.38629  1.38629  1.38629  1.38629
          0.06034  3.53100  3.53100  1.46634  0.26236  1.09861  0.40547
      2   2.19189  1.97227  2.26310  0.43831      2 - -
          1.38629  1.38629  1.38629  1.38629
          0.06034  3.53100  3.53100  1.46634  0.26236  1.09861  0.40547
      3   2.12548  2.54191  0.34386  2.37646      3 - -
          1.38629  1.38629  1.38629  1.38629
          0.06034  3.53100  3.53100  1.46634  0.26236  1.09861  0.40547
      4   1.82247  2.39915  0.44485  2.23812      4 - -
          1.38629  1.38629  1.38629  1.38629
          0.06034  3.53100  3.53100  1.46634  0.26236  1.09861  0.40547
      5   2.12548  2.54191  0.34386  2.37646      5 - -
          1.38629  1.38629  1.38629  1.38629
          0.06034  3.53100  3.53100  1.46634  0.26236  1.09861  0.40547
      6   2.12548  2.54191  0.34386  2.37646      6 - -
          1.38629  1.38629  1.38629  1.38629
          0.06034  3.53100  3.53100  1.46634  0.26236  1.09861  0.40547
```



the number of match states in the model.

Symbol alphabet type

the number of sequences that the HMM was trained on

the effective total number of sequences determined by hmmbuild during sequence weighting, for combining observed counts with Dirichlet prior information in parameterizing the model

This number is calculated from the training sequence data, and used in conjunction with the alignment map information to verify that a given alignment is indeed the alignment that the map is for

Reference annotation flag;.Reference column annotation is picked up from a Stockholm alignment file's #=GC RF line.

Consensus structure annotation flag; is picked up from a Stockholm file's #=GC SS_cons line

```
HMMER3/b [3.0 | March 2010]
NAME  fasta2.aln
LENG  459
ALPH  DNA
RF    no
CS    no
MAP   yes
DATE  Mon Dec  5 12:34:59 2011
NSEQ  10
EFFN  1.215820
CKSUM 3507035756
STATS LOCAL MSV       -11.0255  0.69846
STATS LOCAL VITERBI   -12.6201  0.69846
STATS LOCAL FORWARD    -5.1306  0.69846
HMM          A        C        G        T
            m->m     m->i     m->d     i->m     i->i     d->m     d->d
  COMPO   1.24690  1.59679  1.28051  1.46032
          1.38629  1.38629  1.38629  1.38629
          0.06034  3.53100  3.53100  1.46634  0.26236  0.00000        *
      1   0.38370  2.41911  2.07966  2.25673      1 - -
          1.38629  1.38629  1.38629  1.38629
          0.06034  3.53100  3.53100  1.46634  0.26236  1.09861  0.40547
      2   2.19189  1.97227  2.26310  0.43831      2 - -
          1.38629  1.38629  1.38629  1.38629
          0.06034  3.53100  3.53100  1.46634  0.26236  1.09861  0.40547
      3   2.12548  2.54191  0.34386  2.37646      3 - -
          1.38629  1.38629  1.38629  1.38629
          0.06034  3.53100  3.53100  1.46634  0.26236  1.09861  0.40547
      4   1.82247  2.39915  0.44485  2.23812      4 - -
          1.38629  1.38629  1.38629  1.38629
          0.06034  3.53100  3.53100  1.46634  0.26236  1.09861  0.40547
      5   2.12548  2.54191  0.34386  2.37646      5 - -
          1.38629  1.38629  1.38629  1.38629
          0.06034  3.53100  3.53100  1.46634  0.26236  1.09861  0.40547
      6   2.12548  2.54191  0.34386  2.37646      6 - -
          1.38629  1.38629  1.38629  1.38629
          0.06034  3.53100  3.53100  1.46634  0.26236  1.09861  0.40547
```

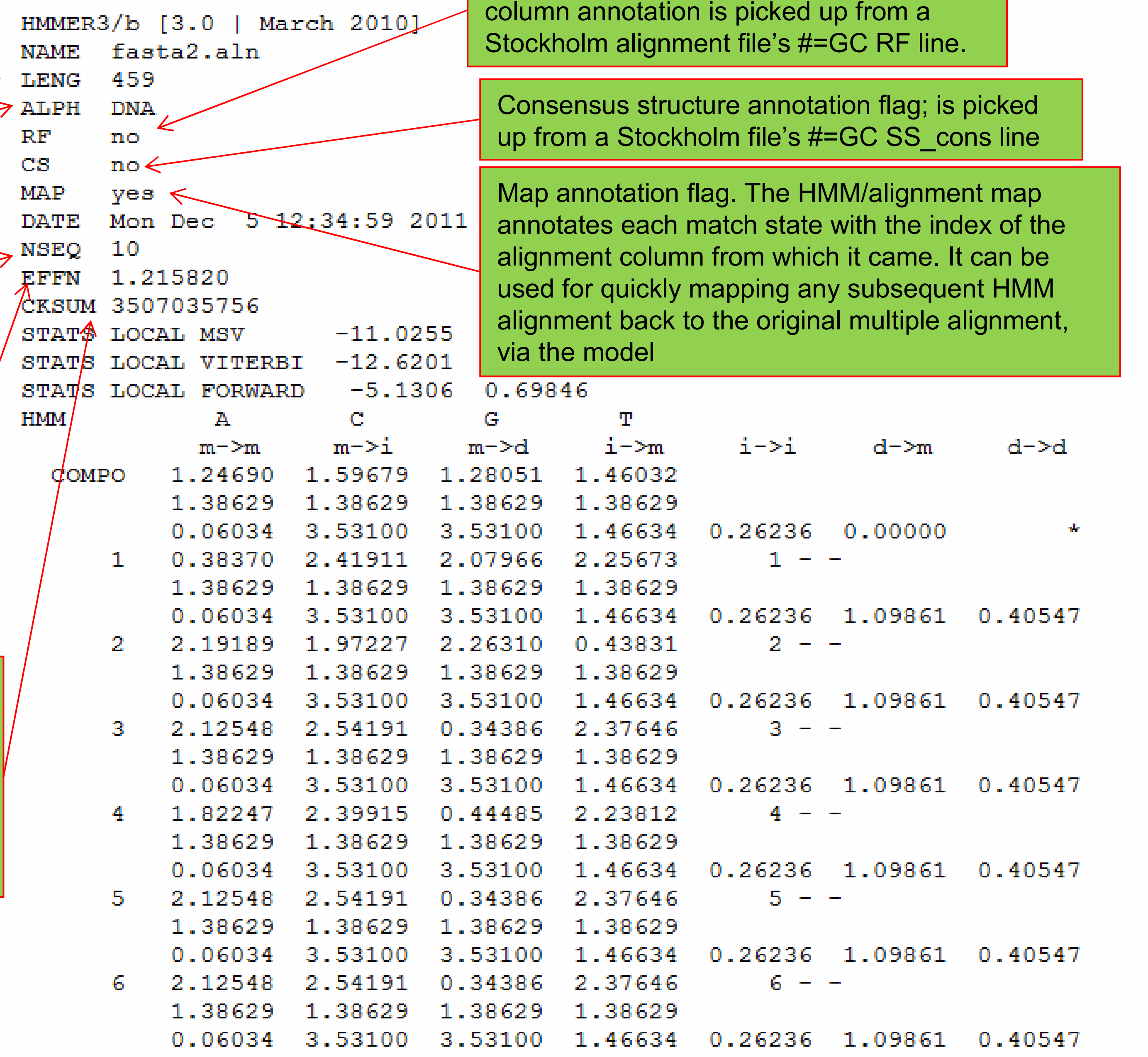
the number of match states in the model.

Symbol alphabet type

the number of sequences that the HMM was trained on

the effective total number of sequences determined by hmmbuild during sequence weighting, for combining observed counts with Dirichlet prior information in parameterizing the model

This number is calculated from the training sequence data, and used in conjunction with the alignment map information to verify that a given alignment is indeed the alignment that the map is for



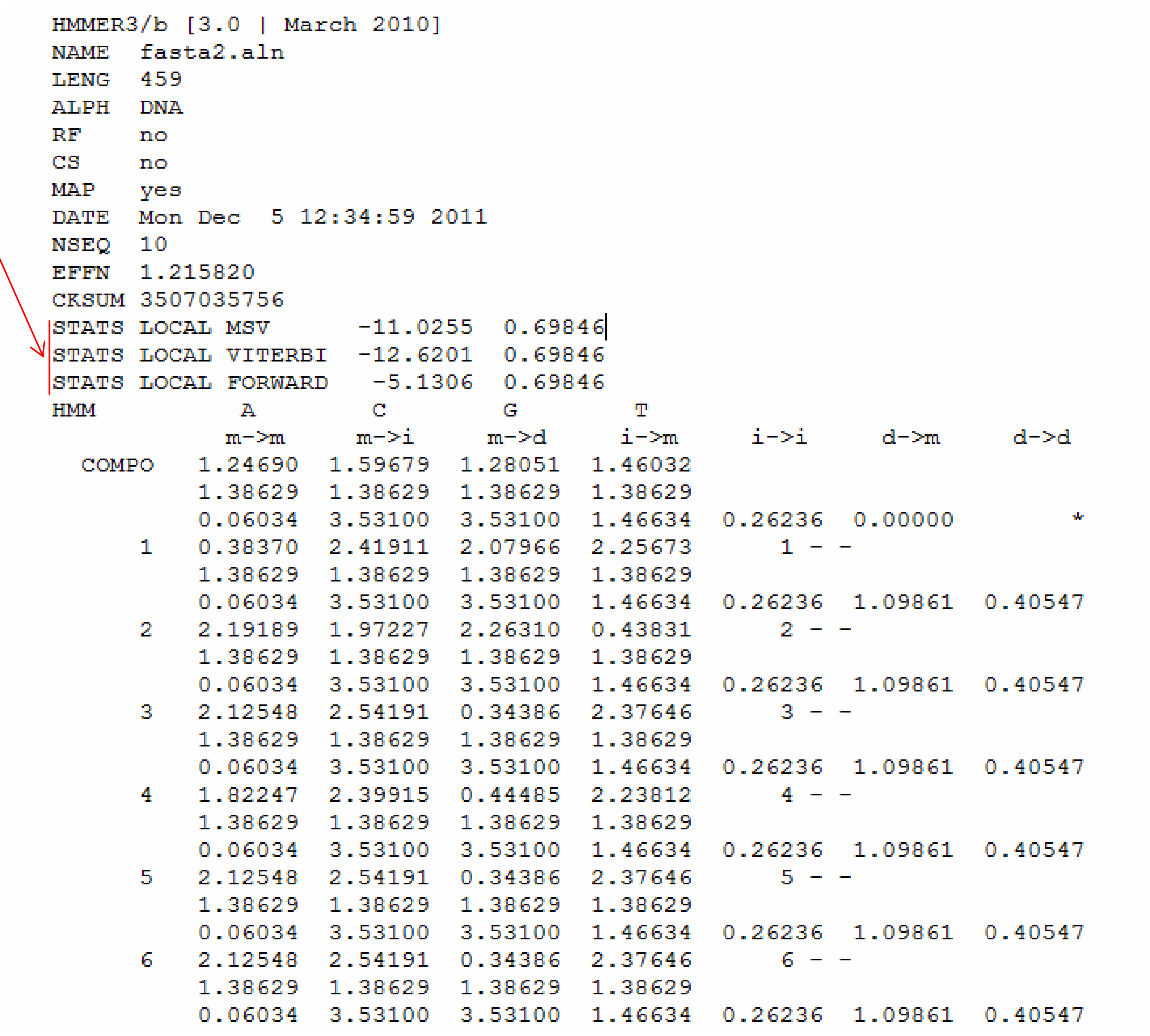
```
HMMER3/b [3.0 | March 2010]
NAME  fasta2.aln
LENG  459
ALPH  DNA
RF    no
CS    no
MAP   yes
DATE  Mon Dec  5 12:34:59 2011
NSEQ  10
EFFN  1.215820
CKSUM 3507035756
STATS LOCAL MSV       -11.0255
STATS LOCAL VITERBI   -12.6201
STATS LOCAL FORWARD    -5.1306  0.69846
HMM          A        C        G        T
            m->m     m->i     m->d     i->m     i->i     d->m     d->d
  COMPO   1.24690  1.59679  1.28051  1.46032
          1.38629  1.38629  1.38629  1.38629
          0.06034  3.53100  3.53100  1.46634  0.26236  0.00000        *
      1   0.38370  2.41911  2.07966  2.25673      1 - -
          1.38629  1.38629  1.38629  1.38629
          0.06034  3.53100  3.53100  1.46634  0.26236  1.09861  0.40547
      2   2.19189  1.97227  2.26310  0.43831      2 - -
          1.38629  1.38629  1.38629  1.38629
          0.06034  3.53100  3.53100  1.46634  0.26236  1.09861  0.40547
      3   2.12548  2.54191  0.34386  2.37646      3 - -
          1.38629  1.38629  1.38629  1.38629
          0.06034  3.53100  3.53100  1.46634  0.26236  1.09861  0.40547
      4   1.82247  2.39915  0.44485  2.23812      4 - -
          1.38629  1.38629  1.38629  1.38629
          0.06034  3.53100  3.53100  1.46634  0.26236  1.09861  0.40547
      5   2.12548  2.54191  0.34386  2.37646      5 - -
          1.38629  1.38629  1.38629  1.38629
          0.06034  3.53100  3.53100  1.46634  0.26236  1.09861  0.40547
      6   2.12548  2.54191  0.34386  2.37646      6 - -
          1.38629  1.38629  1.38629  1.38629
          0.06034  3.53100  3.53100  1.46634  0.26236  1.09861  0.40547
```

Reference annotation flag;.Reference column annotation is picked up from a Stockholm alignment file’s #=GC RF line.

Consensus structure annotation flag; is picked up from a Stockholm file’s #=GC SS_cons line

Map annotation flag. The HMM/alignment map annotates each match state with the index of the alignment column from which it came. It can be used for quickly mapping any subsequent HMM alignment back to the original multiple alignment, via the model

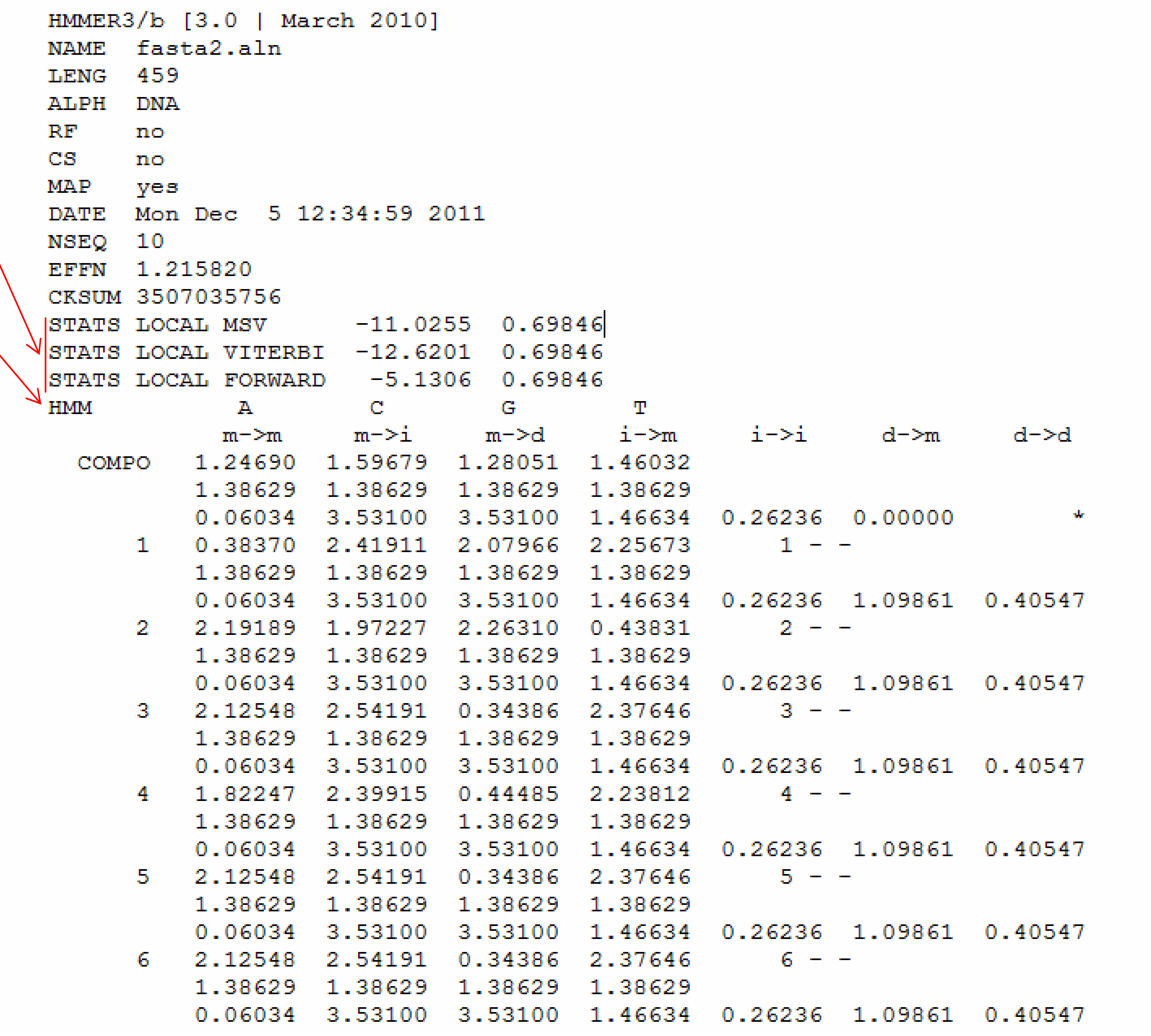
Statistical parameters needed for E-value calculations.



```
HMMER3/b [3.0 | March 2010]
NAME  fasta2.aln
LENG  459
ALPH  DNA
RF    no
CS    no
MAP   yes
DATE  Mon Dec  5 12:34:59 2011
NSEQ  10
EFFN  1.215820
CKSUM 3507035756
STATS LOCAL MSV       -11.0255  0.69846
STATS LOCAL VITERBI   -12.6201  0.69846
STATS LOCAL FORWARD    -5.1306  0.69846
HMM          A        C        G        T
            m->m     m->i     m->d     i->m     i->i     d->m     d->d
  COMPO   1.24690  1.59679  1.28051  1.46032
          1.38629  1.38629  1.38629  1.38629
          0.06034  3.53100  3.53100  1.46634  0.26236  0.00000        *
      1   0.38370  2.41911  2.07966  2.25673      1 - -
          1.38629  1.38629  1.38629  1.38629
          0.06034  3.53100  3.53100  1.46634  0.26236  1.09861  0.40547
      2   2.19189  1.97227  2.26310  0.43831      2 - -
          1.38629  1.38629  1.38629  1.38629
          0.06034  3.53100  3.53100  1.46634  0.26236  1.09861  0.40547
      3   2.12548  2.54191  0.34386  2.37646      3 - -
          1.38629  1.38629  1.38629  1.38629
          0.06034  3.53100  3.53100  1.46634  0.26236  1.09861  0.40547
      4   1.82247  2.39915  0.44485  2.23812      4 - -
          1.38629  1.38629  1.38629  1.38629
          0.06034  3.53100  3.53100  1.46634  0.26236  1.09861  0.40547
      5   2.12548  2.54191  0.34386  2.37646      5 - -
          1.38629  1.38629  1.38629  1.38629
          0.06034  3.53100  3.53100  1.46634  0.26236  1.09861  0.40547
      6   2.12548  2.54191  0.34386  2.37646      6 - -
          1.38629  1.38629  1.38629  1.38629
          0.06034  3.53100  3.53100  1.46634  0.26236  1.09861  0.40547
```

Statistical parameters needed for E-value calculations.

Flags the start of the main model section

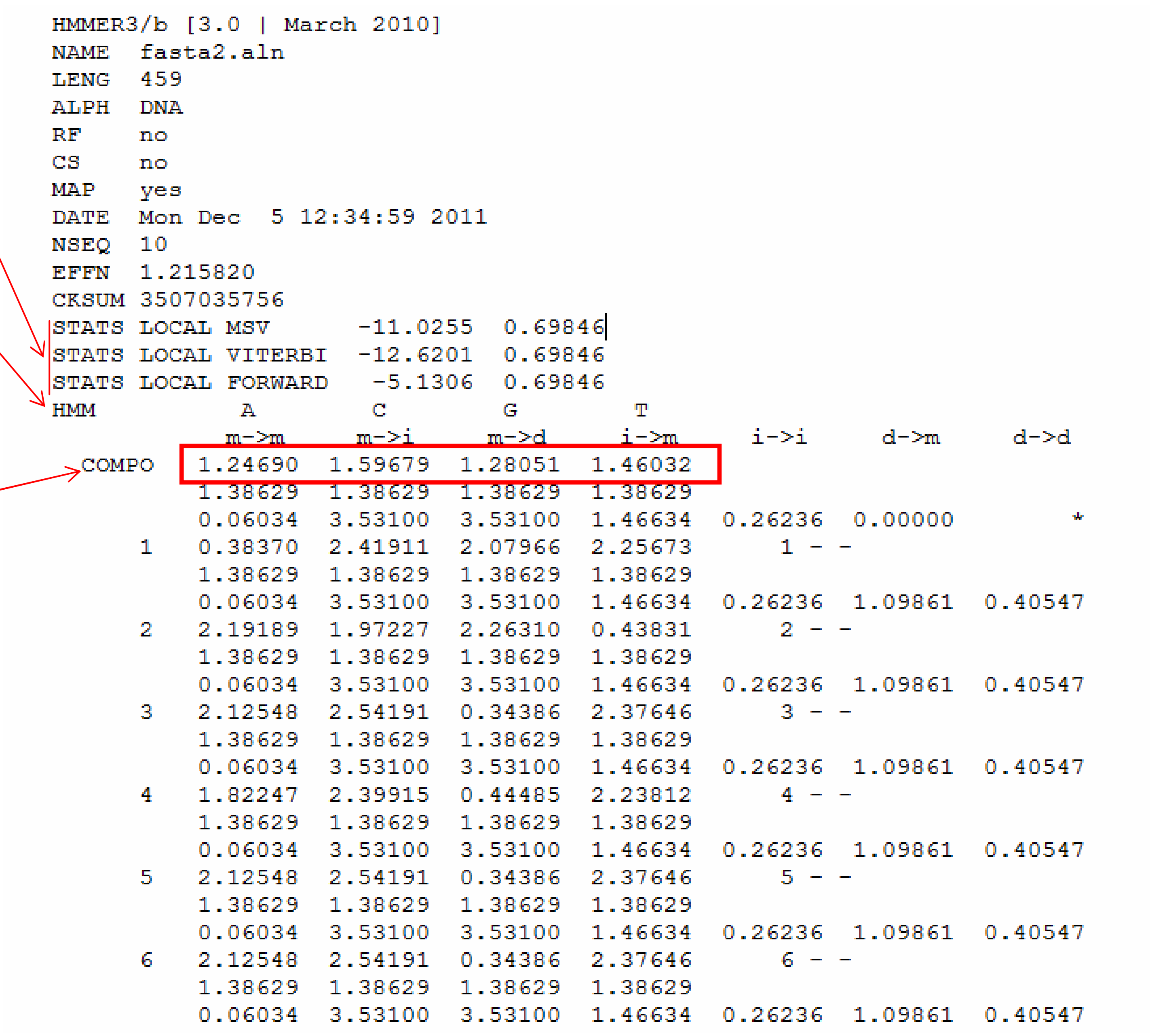


```
HMMER3/b [3.0 | March 2010]
NAME  fasta2.aln
LENG  459
ALPH  DNA
RF    no
CS    no
MAP   yes
DATE  Mon Dec  5 12:34:59 2011
NSEQ  10
EFFN  1.215820
CKSUM 3507035756
STATS LOCAL MSV       -11.0255  0.69846
STATS LOCAL VITERBI   -12.6201  0.69846
STATS LOCAL FORWARD    -5.1306  0.69846
HMM          A        C        G        T
            m->m     m->i     m->d     i->m     i->i     d->m     d->d
  COMPO   1.24690  1.59679  1.28051  1.46032
          1.38629  1.38629  1.38629  1.38629
          0.06034  3.53100  3.53100  1.46634  0.26236  0.00000        *
      1   0.38370  2.41911  2.07966  2.25673      1 - -
          1.38629  1.38629  1.38629  1.38629
          0.06034  3.53100  3.53100  1.46634  0.26236  1.09861  0.40547
      2   2.19189  1.97227  2.26310  0.43831      2 - -
          1.38629  1.38629  1.38629  1.38629
          0.06034  3.53100  3.53100  1.46634  0.26236  1.09861  0.40547
      3   2.12548  2.54191  0.34386  2.37646      3 - -
          1.38629  1.38629  1.38629  1.38629
          0.06034  3.53100  3.53100  1.46634  0.26236  1.09861  0.40547
      4   1.82247  2.39915  0.44485  2.23812      4 - -
          1.38629  1.38629  1.38629  1.38629
          0.06034  3.53100  3.53100  1.46634  0.26236  1.09861  0.40547
      5   2.12548  2.54191  0.34386  2.37646      5 - -
          1.38629  1.38629  1.38629  1.38629
          0.06034  3.53100  3.53100  1.46634  0.26236  1.09861  0.40547
      6   2.12548  2.54191  0.34386  2.37646      6 - -
          1.38629  1.38629  1.38629  1.38629
          0.06034  3.53100  3.53100  1.46634  0.26236  1.09861  0.40547
```

Statistical parameters needed for E-value calculations.

Flags the start of the main model section

these are the model’s overall average match state emission probabilities, which are used as a background residue composition in the “filter null” model.



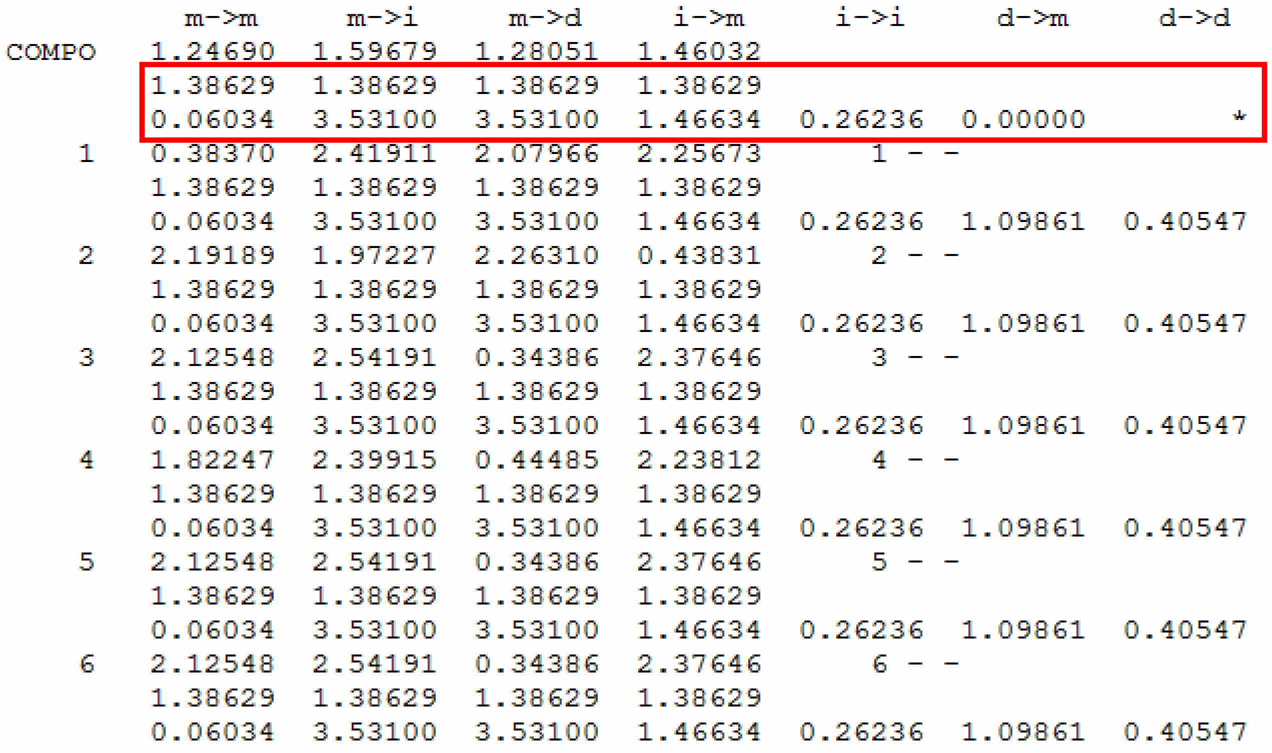
```
HMMER3/b [3.0 | March 2010]
NAME  fasta2.aln
LENG  459
ALPH  DNA
RF    no
CS    no
MAP   yes
DATE  Mon Dec  5 12:34:59 2011
NSEQ  10
EFFN  1.215820
CKSUM 3507035756
STATS LOCAL MSV       -11.0255  0.69846
STATS LOCAL VITERBI   -12.6201  0.69846
STATS LOCAL FORWARD    -5.1306  0.69846
HMM          A        C        G        T
            m->m     m->i     m->d     i->m     i->i     d->m     d->d
  COMPO   1.24690  1.59679  1.28051  1.46032
          1.38629  1.38629  1.38629  1.38629
          0.06034  3.53100  3.53100  1.46634  0.26236  0.00000        *
      1   0.38370  2.41911  2.07966  2.25673      1 - -
          1.38629  1.38629  1.38629  1.38629
          0.06034  3.53100  3.53100  1.46634  0.26236  1.09861  0.40547
      2   2.19189  1.97227  2.26310  0.43831      2 - -
          1.38629  1.38629  1.38629  1.38629
          0.06034  3.53100  3.53100  1.46634  0.26236  1.09861  0.40547
      3   2.12548  2.54191  0.34386  2.37646      3 - -
          1.38629  1.38629  1.38629  1.38629
          0.06034  3.53100  3.53100  1.46634  0.26236  1.09861  0.40547
      4   1.82247  2.39915  0.44485  2.23812      4 - -
          1.38629  1.38629  1.38629  1.38629
          0.06034  3.53100  3.53100  1.46634  0.26236  1.09861  0.40547
      5   2.12548  2.54191  0.34386  2.37646      5 - -
          1.38629  1.38629  1.38629  1.38629
          0.06034  3.53100  3.53100  1.46634  0.26236  1.09861  0.40547
      6   2.12548  2.54191  0.34386  2.37646      6 - -
          1.38629  1.38629  1.38629  1.38629
          0.06034  3.53100  3.53100  1.46634  0.26236  1.09861  0.40547
```

contains information for the core model’s BEGIN node. This is stored as model node 0, and match state 0 is treated as the BEGIN state.

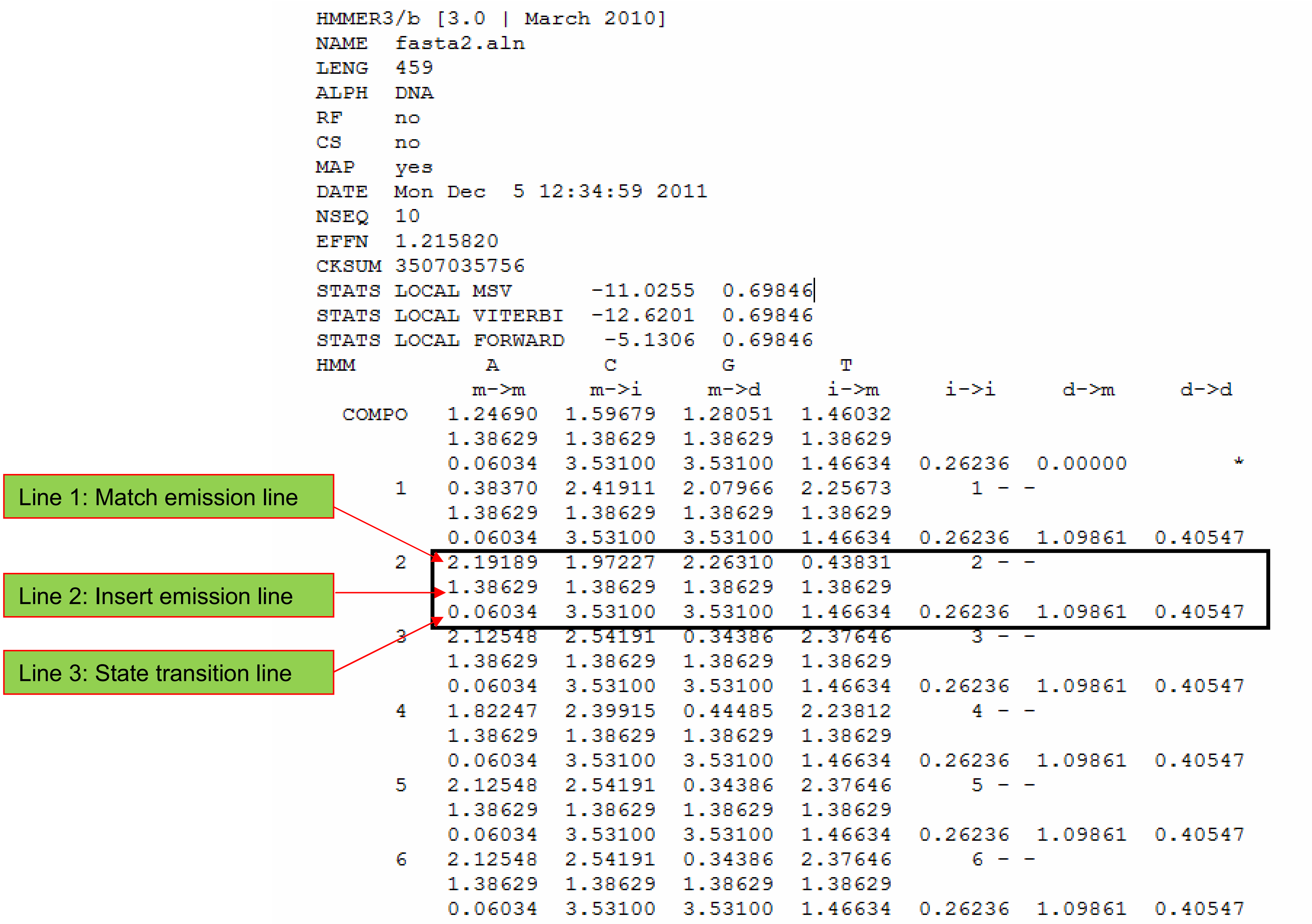
The begin state is mute, so there are no match emission probabilities. The first line is the insert 0 emissions. The second line contains the transitions from the begin state and insert state 0. These seven numbers are:

B → M1, B → I_0, B → D1; I_0 → M1, I0 → I_0;

then a 0.0 and a ’*’, because by convention, nonexistent transitions from the nonexistent delete state 0 are set to log 1 = 0 and log 0 = -∞ = ‘*’.



```
              m->m     m->i     m->d     i->m     i->i     d->m     d->d
  COMPO     1.24690  1.59679  1.28051  1.46032
            1.38629  1.38629  1.38629  1.38629
            0.06034  3.53100  3.53100  1.46634  0.26236  0.00000        *
       1    0.38370  2.41911  2.07966  2.25673        1 - -
            1.38629  1.38629  1.38629  1.38629
            0.06034  3.53100  3.53100  1.46634  0.26236  1.09861  0.40547
       2    2.19189  1.97227  2.26310  0.43831        2 - -
            1.38629  1.38629  1.38629  1.38629
            0.06034  3.53100  3.53100  1.46634  0.26236  1.09861  0.40547
       3    2.12548  2.54191  0.34386  2.37646        3 - -
            1.38629  1.38629  1.38629  1.38629
            0.06034  3.53100  3.53100  1.46634  0.26236  1.09861  0.40547
       4    1.82247  2.39915  0.44485  2.23812        4 - -
            1.38629  1.38629  1.38629  1.38629
            0.06034  3.53100  3.53100  1.46634  0.26236  1.09861  0.40547
       5    2.12548  2.54191  0.34386  2.37646        5 - -
            1.38629  1.38629  1.38629  1.38629
            0.06034  3.53100  3.53100  1.46634  0.26236  1.09861  0.40547
       6    2.12548  2.54191  0.34386  2.37646        6 - -
            1.38629  1.38629  1.38629  1.38629
            0.06034  3.53100  3.53100  1.46634  0.26236  1.09861  0.40547
```

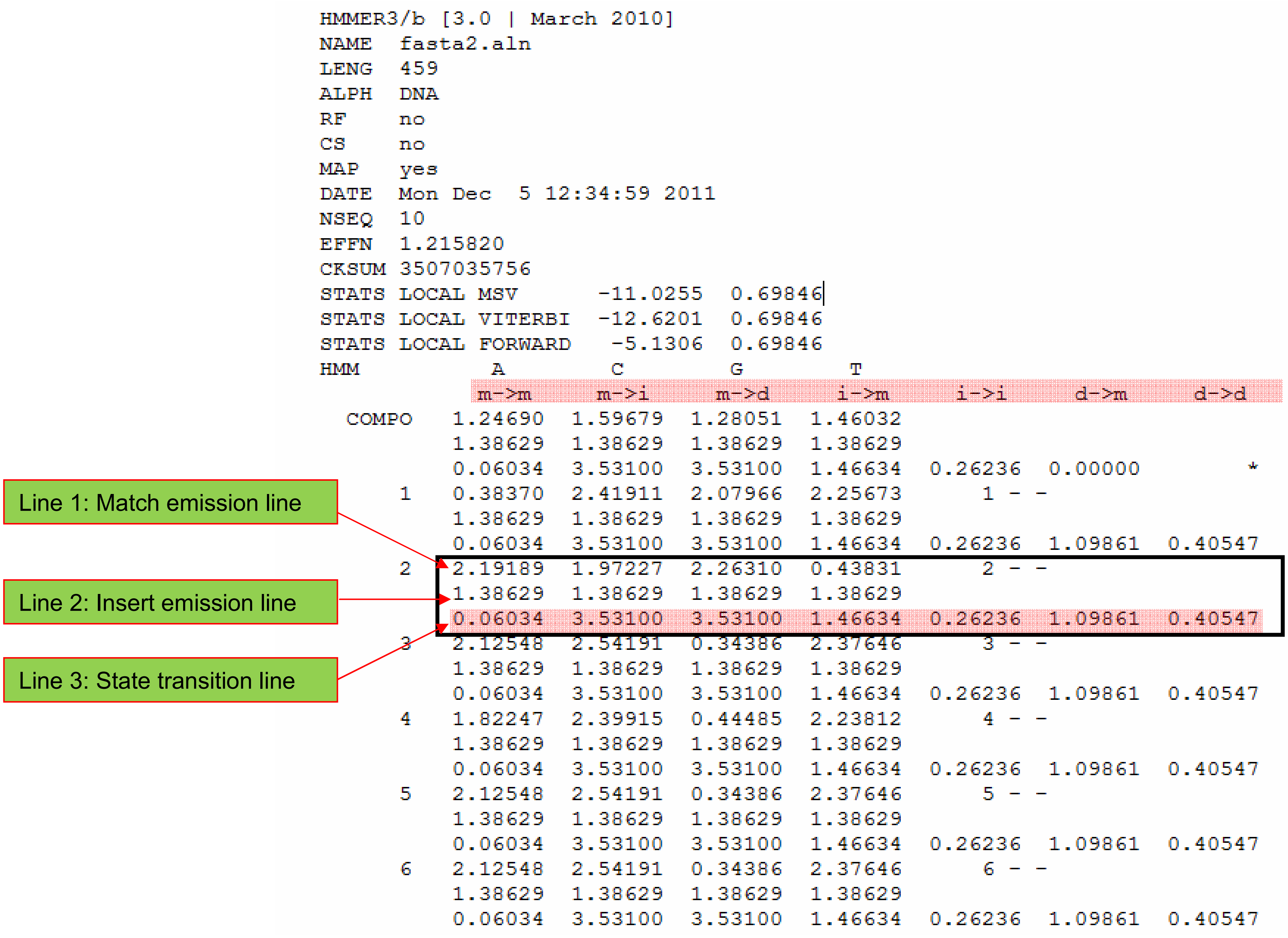


```
HMMER3/b [3.0 | March 2010]
NAME  fasta2.aln
LENG  459
ALPH  DNA
RF    no
CS    no
MAP   yes
DATE  Mon Dec  5 12:34:59 2011
NSEQ  10
EFFN  1.215820
CKSUM 3507035756
STATS LOCAL MSV       -11.0255  0.69846
STATS LOCAL VITERBI   -12.6201  0.69846
STATS LOCAL FORWARD    -5.1306  0.69846
HMM          A        C        G        T
           m->m     m->i     m->d     i->m     i->i     d->m     d->d
  COMPO   1.24690  1.59679  1.28051  1.46032
          1.38629  1.38629  1.38629  1.38629
          0.06034  3.53100  3.53100  1.46634  0.26236  0.00000        *
      1   0.38370  2.41911  2.07966  2.25673      1 - -
          1.38629  1.38629  1.38629  1.38629
          0.06034  3.53100  3.53100  1.46634  0.26236  1.09861  0.40547
      2   2.19189  1.97227  2.26310  0.43831      2 - -
          1.38629  1.38629  1.38629  1.38629
          0.06034  3.53100  3.53100  1.46634  0.26236  1.09861  0.40547
      3   2.12548  2.54191  0.34386  2.37646      3 - -
          1.38629  1.38629  1.38629  1.38629
          0.06034  3.53100  3.53100  1.46634  0.26236  1.09861  0.40547
      4   1.82247  2.39915  0.44485  2.23812      4 - -
          1.38629  1.38629  1.38629  1.38629
          0.06034  3.53100  3.53100  1.46634  0.26236  1.09861  0.40547
      5   2.12548  2.54191  0.34386  2.37646      5 - -
          1.38629  1.38629  1.38629  1.38629
          0.06034  3.53100  3.53100  1.46634  0.26236  1.09861  0.40547
      6   2.12548  2.54191  0.34386  2.37646      6 - -
          1.38629  1.38629  1.38629  1.38629
          0.06034  3.53100  3.53100  1.46634  0.26236  1.09861  0.40547
```

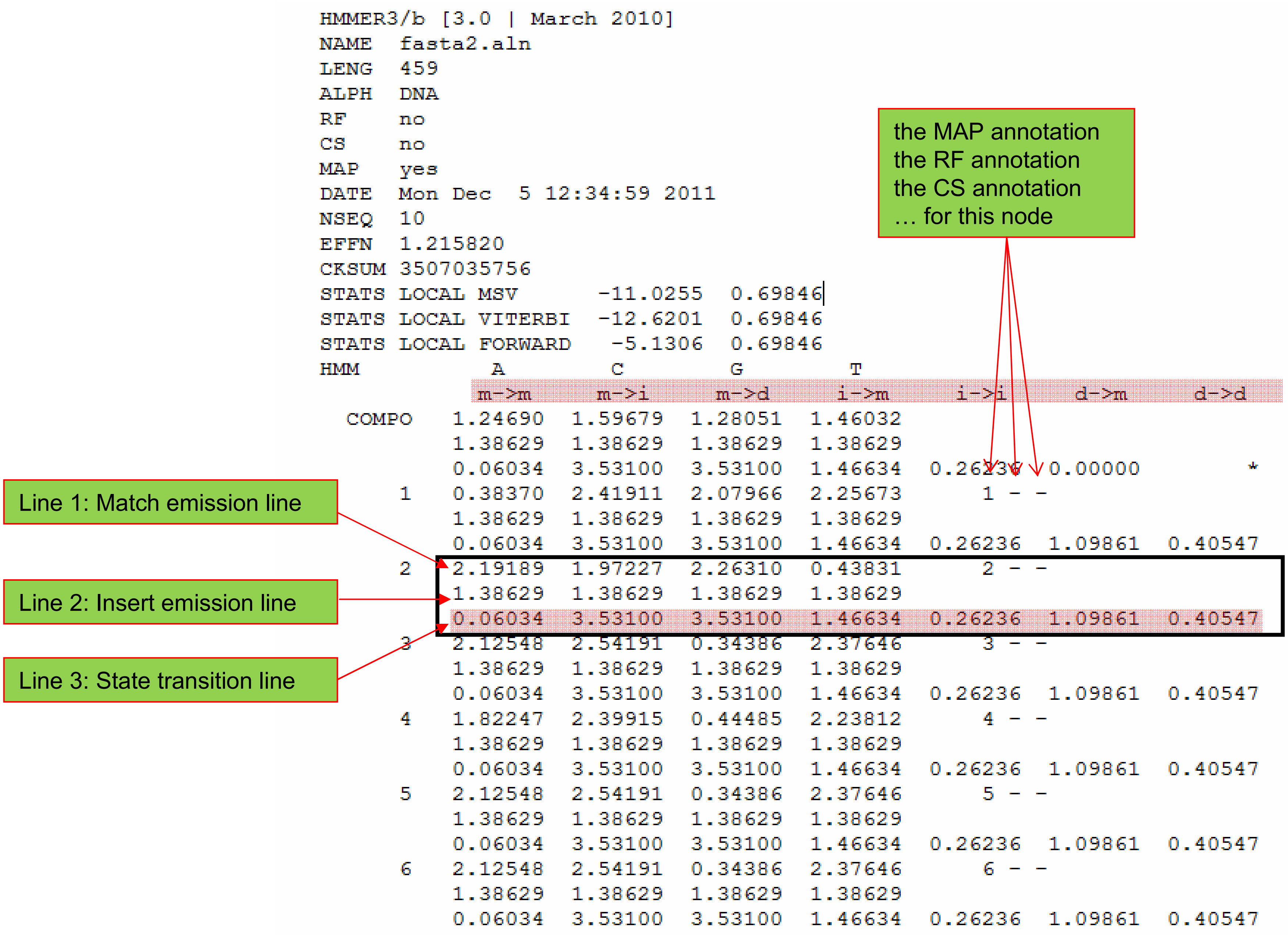
Line 1: Match emission line

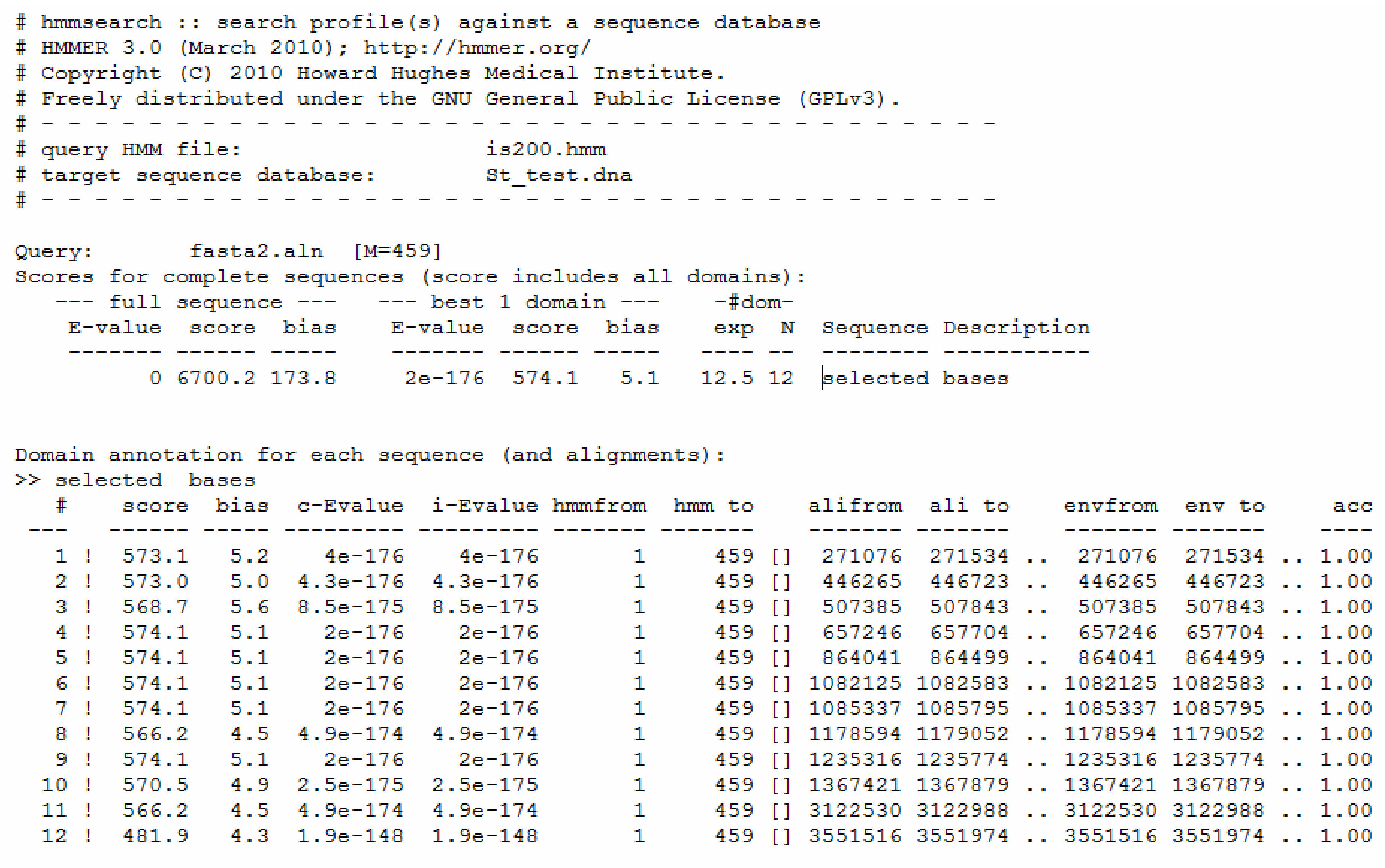
Line 2: Insert emission line

Line 3: State transition line



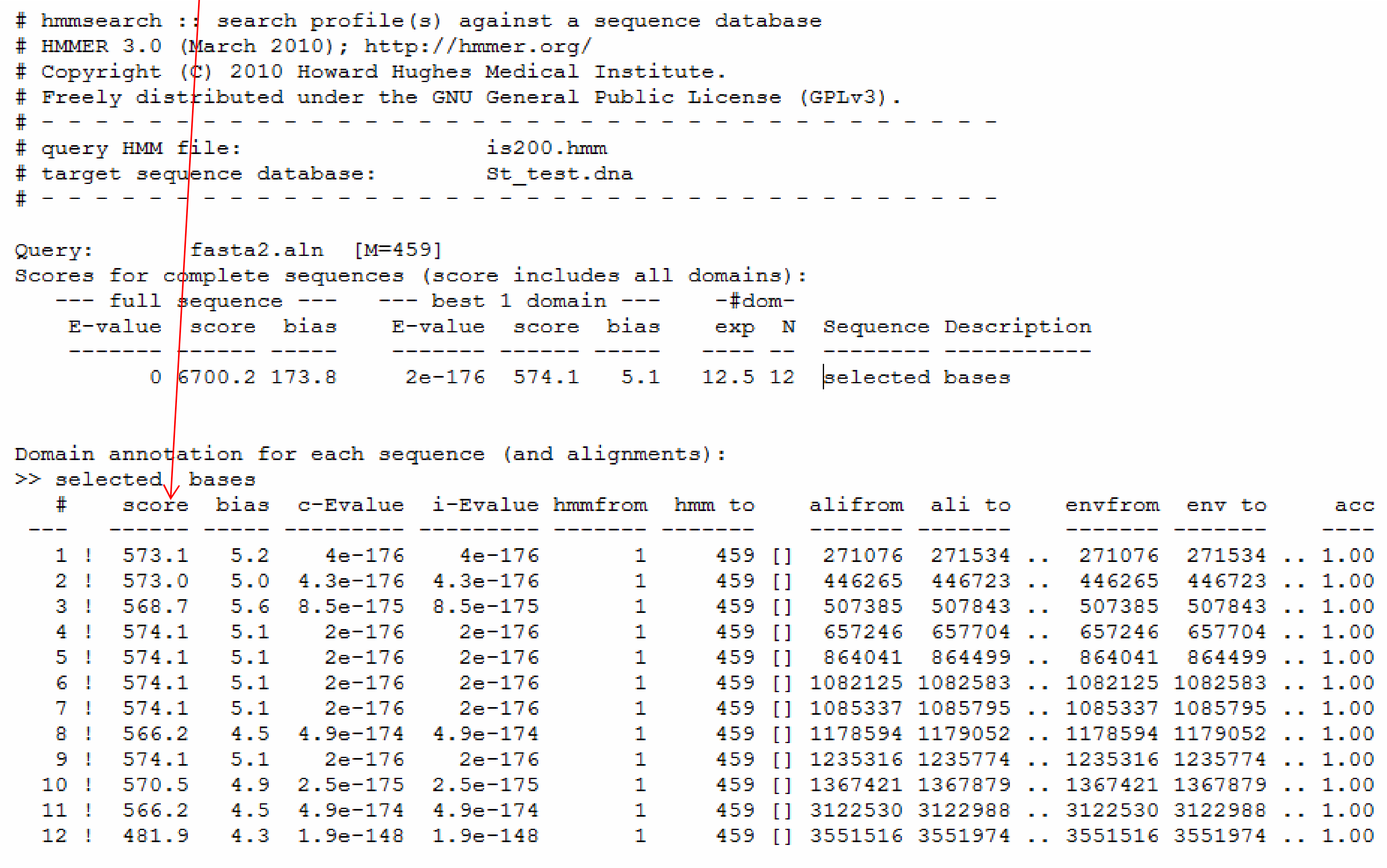
```
HMMER3/b [3.0 | March 2010]
NAME  fasta2.aln
LENG  459
ALPH  DNA
RF    no
CS    no
MAP   yes
DATE  Mon Dec  5 12:34:59 2011
NSEQ  10
EFFN  1.215820
CKSUM 3507035756
STATS LOCAL MSV       -11.0255  0.69846
STATS LOCAL VITERBI   -12.6201  0.69846
STATS LOCAL FORWARD    -5.1306  0.69846
HMM          A        C        G        T
            m->m     m->i     m->d     i->m     i->i     d->m     d->d
  COMPO   1.24690  1.59679  1.28051  1.46032
          1.38629  1.38629  1.38629  1.38629
          0.06034  3.53100  3.53100  1.46634  0.26236  0.00000        *
      1   0.38370  2.41911  2.07966  2.25673      1 - -
          1.38629  1.38629  1.38629  1.38629
          0.06034  3.53100  3.53100  1.46634  0.26236  1.09861  0.40547
      2   2.19189  1.97227  2.26310  0.43831      2 - -
          1.38629  1.38629  1.38629  1.38629
          0.06034  3.53100  3.53100  1.46634  0.26236  1.09861  0.40547
      3   2.12548  2.54191  0.34386  2.37646      3 - -
          1.38629  1.38629  1.38629  1.38629
          0.06034  3.53100  3.53100  1.46634  0.26236  1.09861  0.40547
      4   1.82247  2.39915  0.44485  2.23812      4 - -
          1.38629  1.38629  1.38629  1.38629
          0.06034  3.53100  3.53100  1.46634  0.26236  1.09861  0.40547
      5   2.12548  2.54191  0.34386  2.37646      5 - -
          1.38629  1.38629  1.38629  1.38629
          0.06034  3.53100  3.53100  1.46634  0.26236  1.09861  0.40547
      6   2.12548  2.54191  0.34386  2.37646      6 - -
          1.38629  1.38629  1.38629  1.38629
          0.06034  3.53100  3.53100  1.46634  0.26236  1.09861  0.40547
```





```
# hmmsearch :: search profile(s) against a sequence database
# HMMER 3.0 (March 2010); http://hmmer.org/
# Copyright (C) 2010 Howard Hughes Medical Institute.
# Freely distributed under the GNU General Public License (GPLv3).
# - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - -
# query HMM file:                  is200.hmm
# target sequence database:        St_test.dna
# - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - -

Query:       fasta2.aln  [M=459]
Scores for complete sequences (score includes all domains):
   --- full sequence ---   --- best 1 domain ---    -#dom-
    E-value  score  bias    E-value  score  bias    exp  N  Sequence Description
    ------- ------ -----    ------- ------ -----   ---- --  -------- -----------
          0 6700.2 173.8     2e-176  574.1   5.1   12.5 12  selected bases


Domain annotation for each sequence (and alignments):
>> selected  bases
   #    score  bias  c-Evalue  i-Evalue hmmfrom  hmm to    alifrom  ali to    envfrom  env to     acc
 ---   ------ ----- --------- --------- ------- -------    ------- -------    ------- -------    ----
   1 !  573.1   5.2    4e-176    4e-176       1     459 []  271076  271534 ..  271076  271534 .. 1.00
   2 !  573.0   5.0  4.3e-176  4.3e-176       1     459 []  446265  446723 ..  446265  446723 .. 1.00
   3 !  568.7   5.6  8.5e-175  8.5e-175       1     459 []  507385  507843 ..  507385  507843 .. 1.00
   4 !  574.1   5.1    2e-176    2e-176       1     459 []  657246  657704 ..  657246  657704 .. 1.00
   5 !  574.1   5.1    2e-176    2e-176       1     459 []  864041  864499 ..  864041  864499 .. 1.00
   6 !  574.1   5.1    2e-176    2e-176       1     459 [] 1082125 1082583 .. 1082125 1082583 .. 1.00
   7 !  574.1   5.1    2e-176    2e-176       1     459 [] 1085337 1085795 .. 1085337 1085795 .. 1.00
   8 !  566.2   4.5  4.9e-174  4.9e-174       1     459 [] 1178594 1179052 .. 1178594 1179052 .. 1.00
   9 !  574.1   5.1    2e-176    2e-176       1     459 [] 1235316 1235774 .. 1235316 1235774 .. 1.00
  10 !  570.5   4.9  2.5e-175  2.5e-175       1     459 [] 1367421 1367879 .. 1367421 1367879 .. 1.00
  11 !  566.2   4.5  4.9e-174  4.9e-174       1     459 [] 3122530 3122988 .. 3122530 3122988 .. 1.00
  12 !  481.9   4.3  1.9e-148  1.9e-148       1     459 [] 3551516 3551974 .. 3551516 3551974 .. 1.00
```

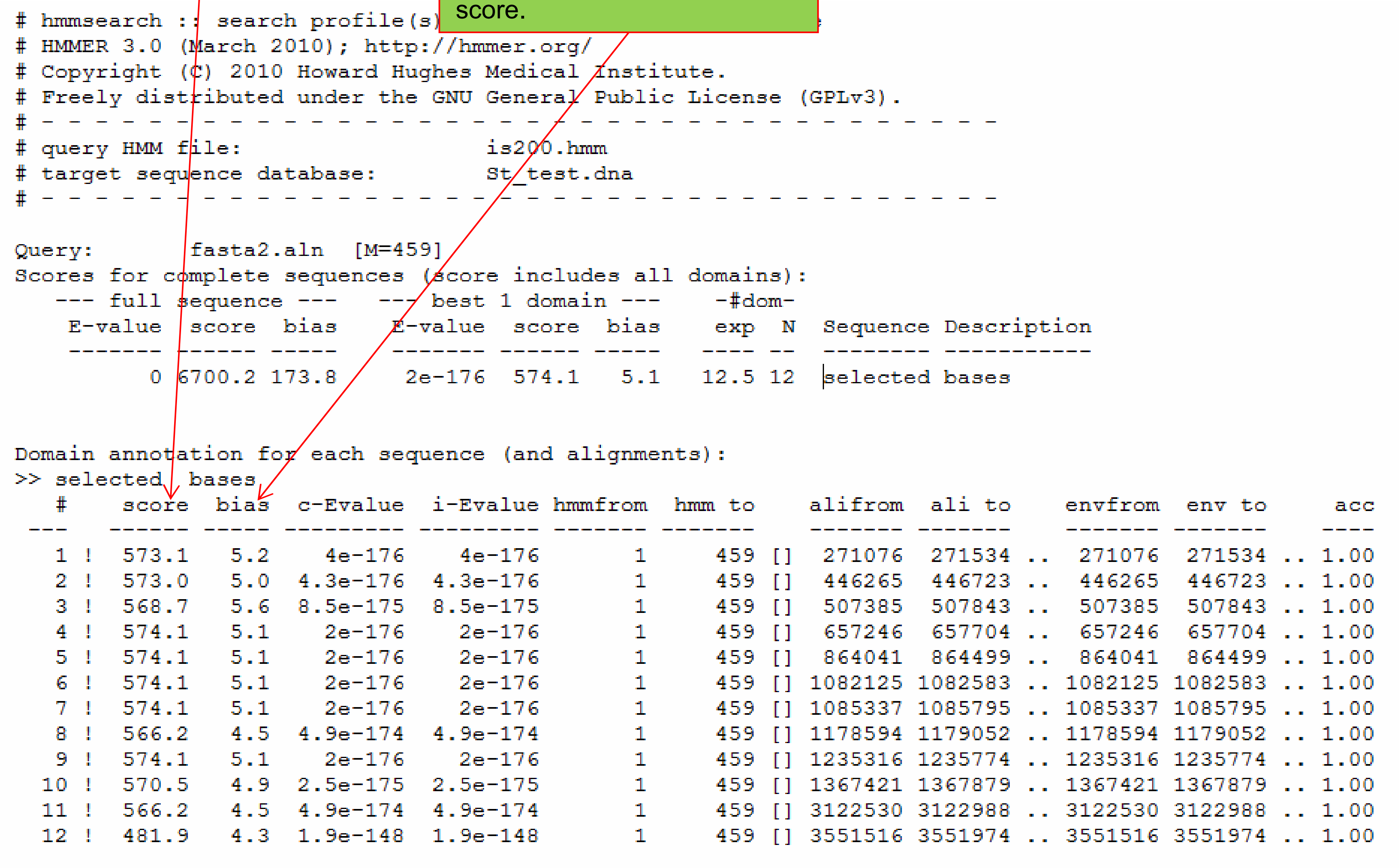
score: The bit score for this domain.



```
# hmmsearch :: search profile(s) against a sequence database
# HMMER 3.0 (March 2010); http://hmmer.org/
# Copyright (C) 2010 Howard Hughes Medical Institute.
# Freely distributed under the GNU General Public License (GPLv3).
# - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - -
# query HMM file:                  is200.hmm
# target sequence database:        St_test.dna
# - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - -

Query:       fasta2.aln  [M=459]
Scores for complete sequences (score includes all domains):
   --- full sequence ---   --- best 1 domain ---    -#dom-
    E-value  score  bias    E-value  score  bias    exp  N  Sequence Description
    ------- ------ -----    ------- ------ -----   ---- --  -------- -----------
          0 6700.2 173.8     2e-176  574.1    5.1   12.5 12  selected bases


Domain annotation for each sequence (and alignments):
>> selected  bases
   #    score  bias  c-Evalue  i-Evalue hmmfrom  hmm to    alifrom  ali to    envfrom  env to     acc
 ---   ------ ----- --------- --------- ------- -------    ------- -------    ------- -------    ----
   1 !  573.1   5.2    4e-176    4e-176       1     459 []  271076  271534 ..  271076  271534 .. 1.00
   2 !  573.0   5.0  4.3e-176  4.3e-176       1     459 []  446265  446723 ..  446265  446723 .. 1.00
   3 !  568.7   5.6  8.5e-175  8.5e-175       1     459 []  507385  507843 ..  507385  507843 .. 1.00
   4 !  574.1   5.1    2e-176    2e-176       1     459 []  657246  657704 ..  657246  657704 .. 1.00
   5 !  574.1   5.1    2e-176    2e-176       1     459 []  864041  864499 ..  864041  864499 .. 1.00
   6 !  574.1   5.1    2e-176    2e-176       1     459 [] 1082125 1082583 .. 1082125 1082583 .. 1.00
   7 !  574.1   5.1    2e-176    2e-176       1     459 [] 1085337 1085795 .. 1085337 1085795 .. 1.00
   8 !  566.2   4.5  4.9e-174  4.9e-174       1     459 [] 1178594 1179052 .. 1178594 1179052 .. 1.00
   9 !  574.1   5.1    2e-176    2e-176       1     459 [] 1235316 1235774 .. 1235316 1235774 .. 1.00
  10 !  570.5   4.9  2.5e-175  2.5e-175       1     459 [] 1367421 1367879 .. 1367421 1367879 .. 1.00
  11 !  566.2   4.5  4.9e-174  4.9e-174       1     459 [] 3122530 3122988 .. 3122530 3122988 .. 1.00
  12 !  481.9   4.3  1.9e-148  1.9e-148       1     459 [] 3551516 3551974 .. 3551516 3551974 .. 1.00
```

score: The bit score for this domain.

The biased composition (null2) score correction that was applied to the domain bit score.

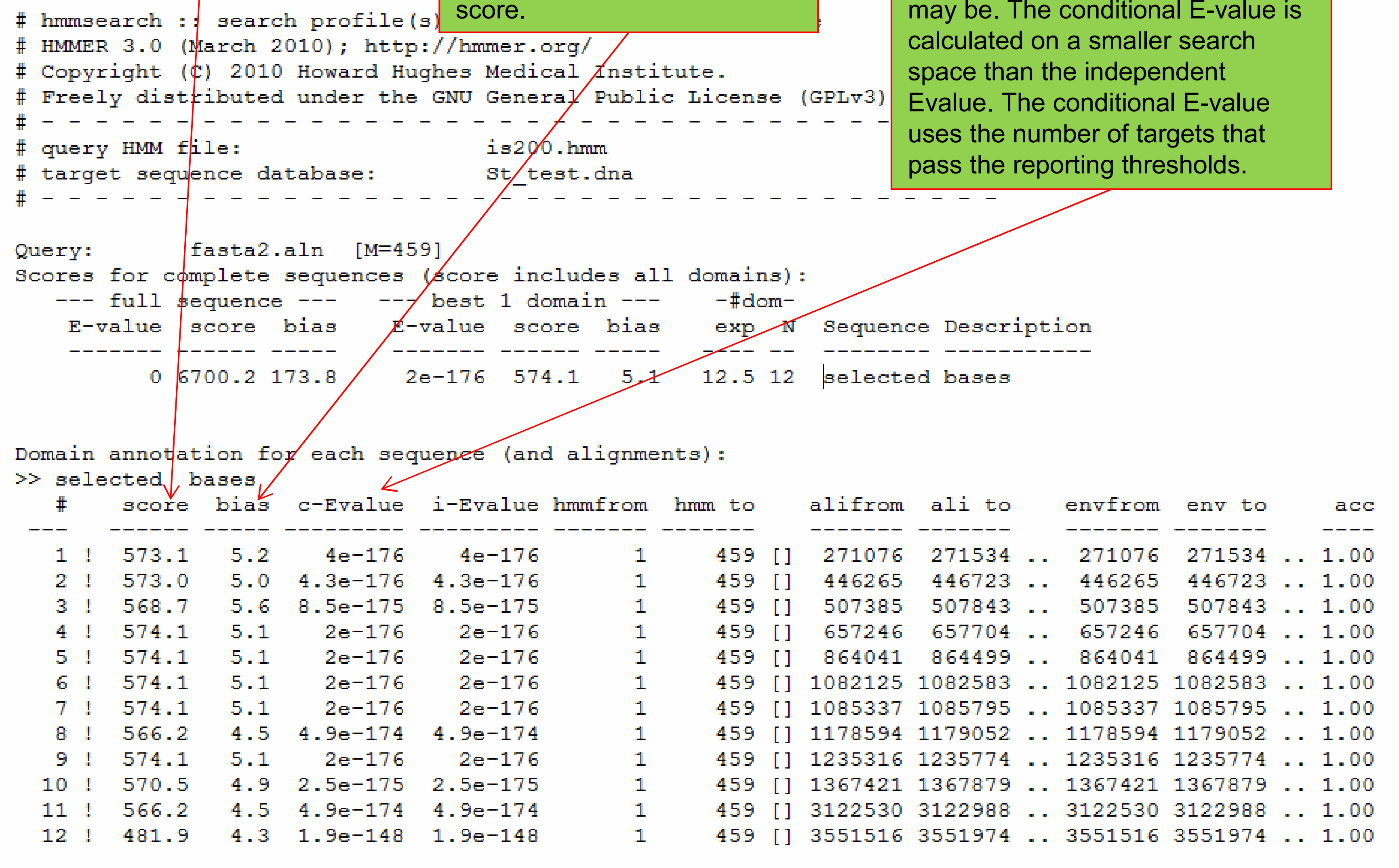


```
# hmmsearch :: search profile(s)
# HMMER 3.0 (March 2010); http://hmmer.org/
# Copyright (C) 2010 Howard Hughes Medical Institute.
# Freely distributed under the GNU General Public License (GPLv3).
# - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - -
# query HMM file:                  is200.hmm
# target sequence database:        St_test.dna
# - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - -

Query:       fasta2.aln  [M=459]
Scores for complete sequences (score includes all domains):
   --- full sequence ---   --- best 1 domain ---    -#dom-
    E-value  score  bias    E-value  score  bias    exp  N  Sequence Description
    ------- ------ -----    ------- ------ -----   ---- --  -------- -----------
          0 6700.2 173.8     2e-176  574.1    5.1   12.5 12  selected bases


Domain annotation for each sequence (and alignments):
>> selected  bases
   #    score  bias  c-Evalue  i-Evalue hmmfrom  hmm to    alifrom  ali to    envfrom  env to     acc
 ---   ------ ----- --------- --------- ------- -------    ------- -------    ------- -------    ----
   1 !  573.1   5.2    4e-176    4e-176       1     459 []  271076  271534 ..  271076  271534 .. 1.00
   2 !  573.0   5.0  4.3e-176  4.3e-176       1     459 []  446265  446723 ..  446265  446723 .. 1.00
   3 !  568.7   5.6  8.5e-175  8.5e-175       1     459 []  507385  507843 ..  507385  507843 .. 1.00
   4 !  574.1   5.1    2e-176    2e-176       1     459 []  657246  657704 ..  657246  657704 .. 1.00
   5 !  574.1   5.1    2e-176    2e-176       1     459 []  864041  864499 ..  864041  864499 .. 1.00
   6 !  574.1   5.1    2e-176    2e-176       1     459 [] 1082125 1082583 .. 1082125 1082583 .. 1.00
   7 !  574.1   5.1    2e-176    2e-176       1     459 [] 1085337 1085795 .. 1085337 1085795 .. 1.00
   8 !  566.2   4.5  4.9e-174  4.9e-174       1     459 [] 1178594 1179052 .. 1178594 1179052 .. 1.00
   9 !  574.1   5.1    2e-176    2e-176       1     459 [] 1235316 1235774 .. 1235316 1235774 .. 1.00
  10 !  570.5   4.9  2.5e-175  2.5e-175       1     459 [] 1367421 1367879 .. 1367421 1367879 .. 1.00
  11 !  566.2   4.5  4.9e-174  4.9e-174       1     459 [] 3122530 3122988 .. 3122530 3122988 .. 1.00
  12 !  481.9   4.3  1.9e-148  1.9e-148       1     459 [] 3551516 3551974 .. 3551516 3551974 .. 1.00
```

score: The bit score for this domain.

The biased composition (null2) score correction that was applied to the domain bit score.

The “conditional E-value”, a permissive measure of how reliable this particular domain may be. The conditional E-value is calculated on a smaller search space than the independent Evalue. The conditional E-value uses the number of targets that pass the reporting thresholds.



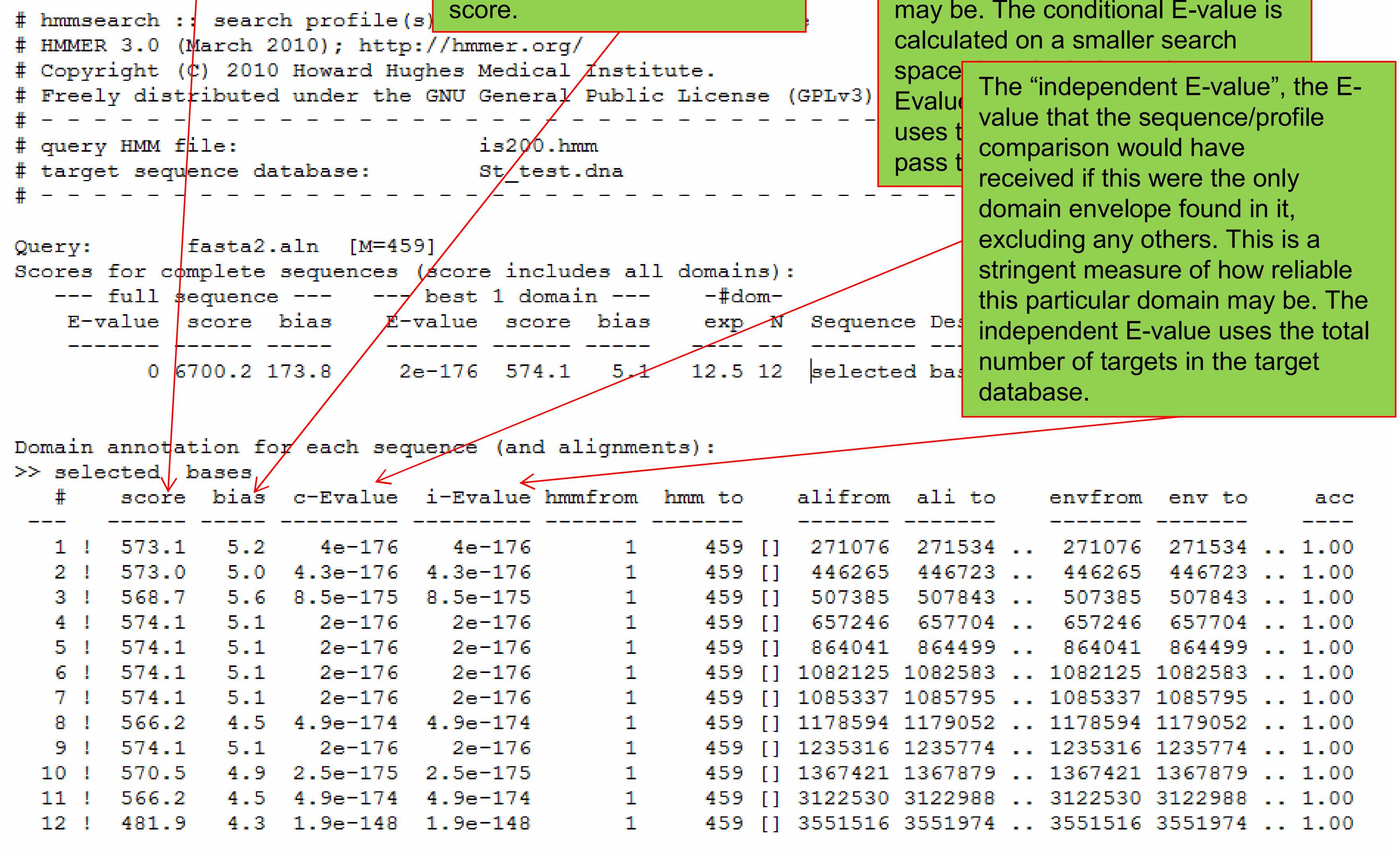
```
# hmmsearch :: search profile(s)
# HMMER 3.0 (March 2010); http://hmmer.org/
# Copyright (C) 2010 Howard Hughes Medical Institute.
# Freely distributed under the GNU General Public License (GPLv3).
# - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - -
# query HMM file:                  is200.hmm
# target sequence database:        St_test.dna
# - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - -

Query:       fasta2.aln  [M=459]
Scores for complete sequences (score includes all domains):
   --- full sequence ---   --- best 1 domain ---   -#dom-
    E-value  score  bias    E-value  score  bias    exp  N  Sequence Description
    ------- ------ -----    ------- ------ -----   ---- --  -------- -----------
          0 6700.2 173.8     2e-176  574.1   5.1   12.5 12  selected bases


Domain annotation for each sequence (and alignments):
>> selected  bases
   #    score  bias  c-Evalue  i-Evalue hmmfrom  hmm to    alifrom  ali to    envfrom  env to     acc
 ---   ------ ----- --------- --------- ------- -------    ------- -------    ------- -------    ----
   1 !  573.1   5.2    4e-176    4e-176       1     459 []  271076  271534 ..  271076  271534 .. 1.00
   2 !  573.0   5.0  4.3e-176  4.3e-176       1     459 []  446265  446723 ..  446265  446723 .. 1.00
   3 !  568.7   5.6  8.5e-175  8.5e-175       1     459 []  507385  507843 ..  507385  507843 .. 1.00
   4 !  574.1   5.1    2e-176    2e-176       1     459 []  657246  657704 ..  657246  657704 .. 1.00
   5 !  574.1   5.1    2e-176    2e-176       1     459 []  864041  864499 ..  864041  864499 .. 1.00
   6 !  574.1   5.1    2e-176    2e-176       1     459 [] 1082125 1082583 .. 1082125 1082583 .. 1.00
   7 !  574.1   5.1    2e-176    2e-176       1     459 [] 1085337 1085795 .. 1085337 1085795 .. 1.00
   8 !  566.2   4.5  4.9e-174  4.9e-174       1     459 [] 1178594 1179052 .. 1178594 1179052 .. 1.00
   9 !  574.1   5.1    2e-176    2e-176       1     459 [] 1235316 1235774 .. 1235316 1235774 .. 1.00
  10 !  570.5   4.9  2.5e-175  2.5e-175       1     459 [] 1367421 1367879 .. 1367421 1367879 .. 1.00
  11 !  566.2   4.5  4.9e-174  4.9e-174       1     459 [] 3122530 3122988 .. 3122530 3122988 .. 1.00
  12 !  481.9   4.3  1.9e-148  1.9e-148       1     459 [] 3551516 3551974 .. 3551516 3551974 .. 1.00
```

score: The bit score for this domain.

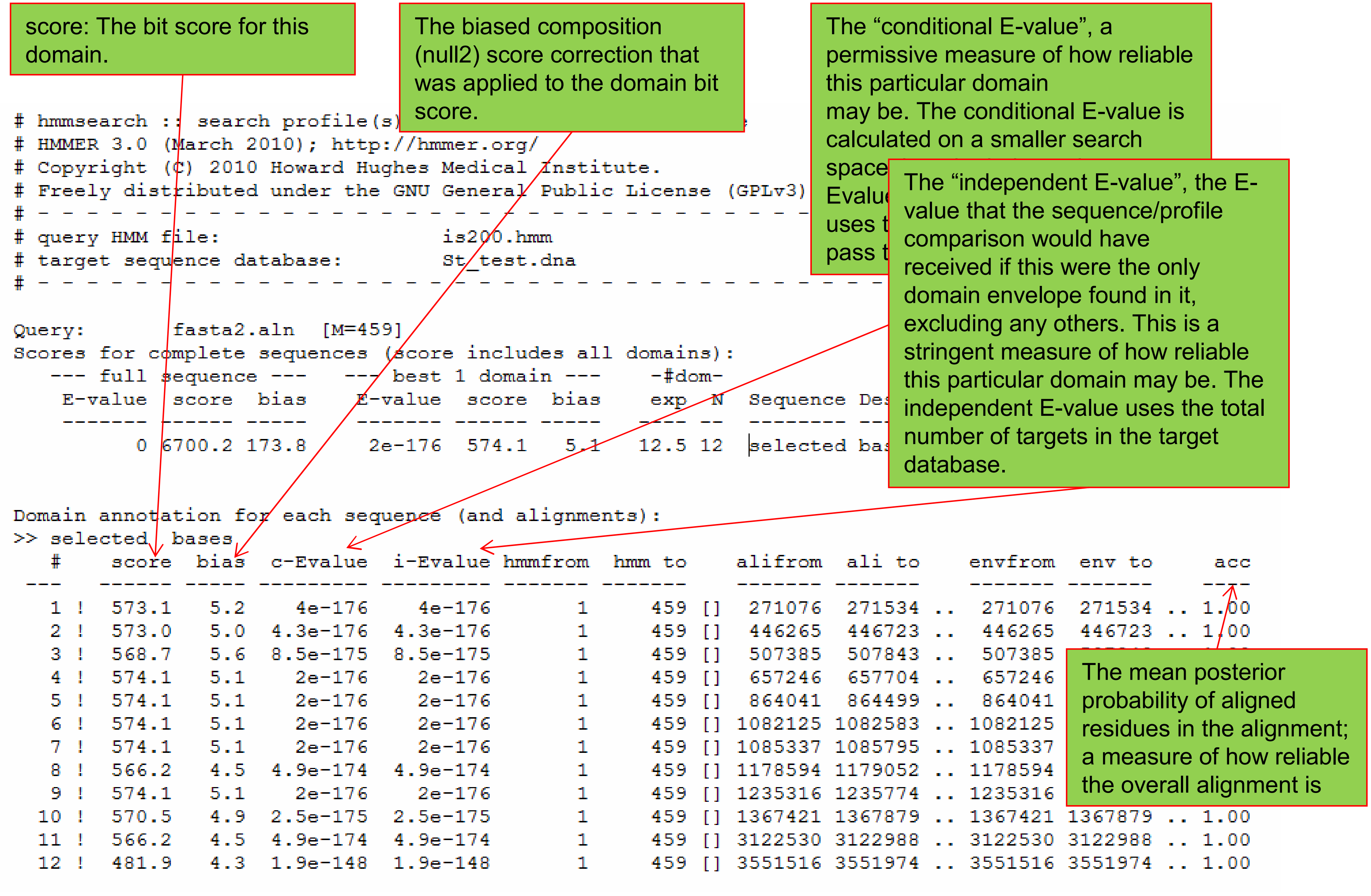
The biased composition (null2) score correction that was applied to the domain bit score.

The “conditional E-value”, a permissive measure of how reliable this particular domain
may be. The conditional E-value is calculated on a smaller search
space
Evalu
uses t
pass t

The “independent E-value”, the E-value that the sequence/profile comparison would have received if this were the only domain envelope found in it, excluding any others. This is a stringent measure of how reliable this particular domain may be. The independent E-value uses the total number of targets in the target database.



```
# hmmsearch :: search profile(s)
# HMMER 3.0 (March 2010); http://hmmer.org/
# Copyright (C) 2010 Howard Hughes Medical Institute.
# Freely distributed under the GNU General Public License (GPLv3)
# - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - -
# query HMM file:                  is200.hmm
# target sequence database:        St_test.dna
# - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - -

Query:       fasta2.aln  [M=459]
Scores for complete sequences (score includes all domains):
   --- full sequence ---    --- best 1 domain ---    -#dom-
    E-value  score  bias    E-value  score  bias    exp  N  Sequence Des
    ------- ------ -----    ------- ------ -----   ---- --  -------- ---
          0 6700.2 173.8      2e-176  574.1   5.1   12.5 12  selected bas


Domain annotation for each sequence (and alignments):
>> selected  bases
   #    score  bias  c-Evalue  i-Evalue hmmfrom  hmm to    alifrom  ali to    envfrom  env to     acc
 ---   ------ ----- --------- --------- ------- -------    ------- -------    ------- -------    ----
   1 !  573.1   5.2    4e-176    4e-176       1     459 []  271076  271534 ..  271076  271534 .. 1.00
   2 !  573.0   5.0  4.3e-176  4.3e-176       1     459 []  446265  446723 ..  446265  446723 .. 1.00
   3 !  568.7   5.6  8.5e-175  8.5e-175       1     459 []  507385  507843 ..  507385  507843 .. 1.00
   4 !  574.1   5.1    2e-176    2e-176       1     459 []  657246  657704 ..  657246  657704 .. 1.00
   5 !  574.1   5.1    2e-176    2e-176       1     459 []  864041  864499 ..  864041  864499 .. 1.00
   6 !  574.1   5.1    2e-176    2e-176       1     459 [] 1082125 1082583 .. 1082125 1082583 .. 1.00
   7 !  574.1   5.1    2e-176    2e-176       1     459 [] 1085337 1085795 .. 1085337 1085795 .. 1.00
   8 !  566.2   4.5  4.9e-174  4.9e-174       1     459 [] 1178594 1179052 .. 1178594 1179052 .. 1.00
   9 !  574.1   5.1    2e-176    2e-176       1     459 [] 1235316 1235774 .. 1235316 1235774 .. 1.00
  10 !  570.5   4.9  2.5e-175  2.5e-175       1     459 [] 1367421 1367879 .. 1367421 1367879 .. 1.00
  11 !  566.2   4.5  4.9e-174  4.9e-174       1     459 [] 3122530 3122988 .. 3122530 3122988 .. 1.00
  12 !  481.9   4.3  1.9e-148  1.9e-148       1     459 [] 3551516 3551974 .. 3551516 3551974 .. 1.00
```



score: The bit score for this domain.

The biased composition (null2) score correction that was applied to the domain bit score.

The “conditional E-value”, a permissive measure of how reliable this particular domain may be. The conditional E-value is calculated on a smaller search space

Evalu

uses t

pass t

The “independent E-value”, the E-value that the sequence/profile comparison would have received if this were the only domain envelope found in it, excluding any others. This is a stringent measure of how reliable this particular domain may be. The independent E-value uses the total number of targets in the target database.

The mean posterior probability of aligned residues in the alignment; a measure of how reliable the overall alignment is

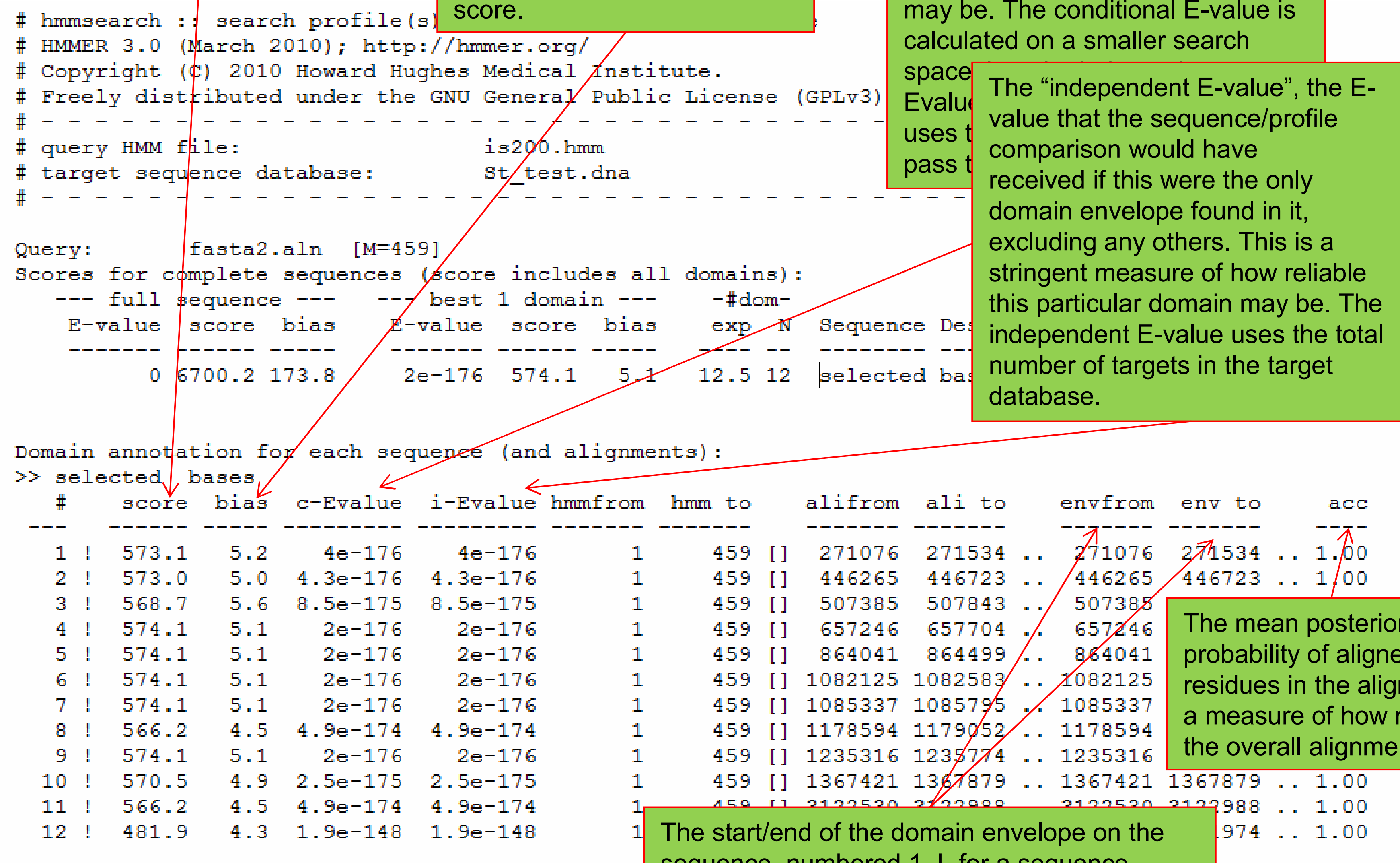
```
# hmmsearch :: search profile(s)
# HMMER 3.0 (March 2010); http://hmmer.org/
# Copyright (C) 2010 Howard Hughes Medical Institute.
# Freely distributed under the GNU General Public License (GPLv3)
# - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - -
# query HMM file:                  is200.hmm
# target sequence database:        St_test.dna
# - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - -

Query:       fasta2.aln  [M=459]
Scores for complete sequences (score includes all domains):
   --- full sequence ---   --- best 1 domain ---    -#dom-
    E-value  score  bias    E-value  score  bias    exp  N  Sequence Des
    ------- ------ -----    ------- ------ -----   ---- --  -------- ---
          0 6700.2 173.8     2e-176  574.1   5.1   12.5 12  selected bas


Domain annotation for each sequence (and alignments):
>> selected  bases
   #    score  bias  c-Evalue  i-Evalue hmmfrom  hmm to    alifrom  ali to    envfrom  env to     acc
 ---   ------ ----- --------- --------- ------- -------    ------- -------    ------- -------    ----
   1 !  573.1   5.2    4e-176    4e-176       1     459 []  271076  271534 ..  271076  271534 .. 1.00
   2 !  573.0   5.0  4.3e-176  4.3e-176       1     459 []  446265  446723 ..  446265  446723 .. 1.00
   3 !  568.7   5.6  8.5e-175  8.5e-175       1     459 []  507385  507843 ..  507385
   4 !  574.1   5.1    2e-176    2e-176       1     459 []  657246  657704 ..  657246
   5 !  574.1   5.1    2e-176    2e-176       1     459 []  864041  864499 ..  864041
   6 !  574.1   5.1    2e-176    2e-176       1     459 [] 1082125 1082583 .. 1082125
   7 !  574.1   5.1    2e-176    2e-176       1     459 [] 1085337 1085795 .. 1085337
   8 !  566.2   4.5  4.9e-174  4.9e-174       1     459 [] 1178594 1179052 .. 1178594
   9 !  574.1   5.1    2e-176    2e-176       1     459 [] 1235316 1235774 .. 1235316
  10 !  570.5   4.9  2.5e-175  2.5e-175       1     459 [] 1367421 1367879 .. 1367421 1367879 .. 1.00
  11 !  566.2   4.5  4.9e-174  4.9e-174       1     459 [] 3122530 3122988 .. 3122530 3122988 .. 1.00
  12 !  481.9   4.3  1.9e-148  1.9e-148       1     459 [] 3551516 3551974 .. 3551516 3551974 .. 1.00
```

score: The bit score for this domain.

The biased composition (null2) score correction that was applied to the domain bit score.

The “conditional E-value”, a permissive measure of how reliable this particular domain may be. The conditional E-value is calculated on a smaller search space
Evalu
uses t
pass t

The “independent E-value”, the E-value that the sequence/profile comparison would have received if this were the only domain envelope found in it, excluding any others. This is a stringent measure of how reliable this particular domain may be. The independent E-value uses the total number of targets in the target database.



```
# hmmsearch :: search profile(s)
# HMMER 3.0 (March 2010); http://hmmer.org/
# Copyright (C) 2010 Howard Hughes Medical Institute.
# Freely distributed under the GNU General Public License (GPLv3)
# - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - -
# query HMM file:                  is200.hmm
# target sequence database:        St_test.dna
# - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - -

Query:       fasta2.aln  [M=459]
Scores for complete sequences (score includes all domains):
   --- full sequence ---   --- best 1 domain ---    -#dom-
    E-value  score  bias    E-value  score  bias    exp  N  Sequence Des
    ------- ------ -----    ------- ------ -----   ---- --  -------- ---
          0 6700.2 173.8     2e-176  574.1   5.1   12.5 12  selected bas


Domain annotation for each sequence (and alignments):
>> selected  bases
   #    score  bias  c-Evalue  i-Evalue hmmfrom  hmm to    alifrom  ali to    envfrom  env to     acc
 ---   ------ ----- --------- --------- ------- -------    ------- -------    ------- -------    ----
   1 !  573.1   5.2    4e-176    4e-176       1     459 []  271076  271534 ..  271076  271534 .. 1.00
   2 !  573.0   5.0  4.3e-176  4.3e-176       1     459 []  446265  446723 ..  446265  446723 .. 1.00
   3 !  568.7   5.6  8.5e-175  8.5e-175       1     459 []  507385  507843 ..  507385
   4 !  574.1   5.1    2e-176    2e-176       1     459 []  657246  657704 ..  657246
   5 !  574.1   5.1    2e-176    2e-176       1     459 []  864041  864499 ..  864041
   6 !  574.1   5.1    2e-176    2e-176       1     459 [] 1082125 1082583 .. 1082125
   7 !  574.1   5.1    2e-176    2e-176       1     459 [] 1085337 1085795 .. 1085337
   8 !  566.2   4.5  4.9e-174  4.9e-174       1     459 [] 1178594 1179052 .. 1178594
   9 !  574.1   5.1    2e-176    2e-176       1     459 [] 1235316 1235774 .. 1235316
  10 !  570.5   4.9  2.5e-175  2.5e-175       1     459 [] 1367421 1367879 .. 1367421 1367879 .. 1.00
  11 !  566.2   4.5  4.9e-174  4.9e-174       1                                3122530 3122988 .. 1.00
  12 !  481.9   4.3  1.9e-148  1.9e-148       1                                            974 .. 1.00
```

The mean posterior probability of aligned residues in the alignment; a measure of how reliable the overall alignment is

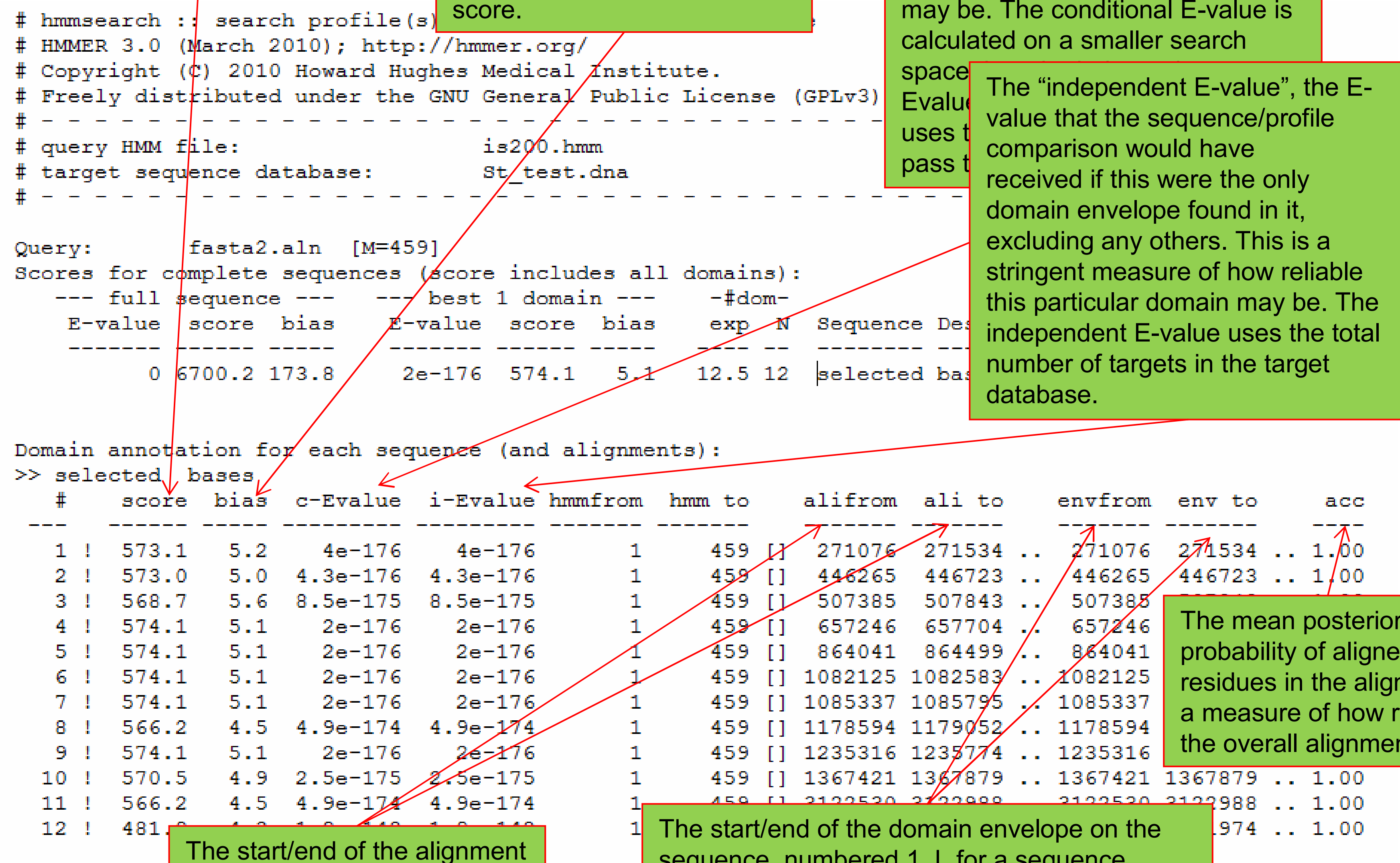
The start/end of the domain envelope on the sequence, numbered 1..L for a sequence of L residues. The envelope defines a subsequence for which their is substantial probability mass supporting a homologous domain

score: The bit score for this domain.

The biased composition (null2) score correction that was applied to the domain bit score.

The "conditional E-value", a permissive measure of how reliable this particular domain may be. The conditional E-value is calculated on a smaller search space
Evalu
uses t
pass t

The "independent E-value", the E-value that the sequence/profile comparison would have received if this were the only domain envelope found in it, excluding any others. This is a stringent measure of how reliable this particular domain may be. The independent E-value uses the total number of targets in the target database.

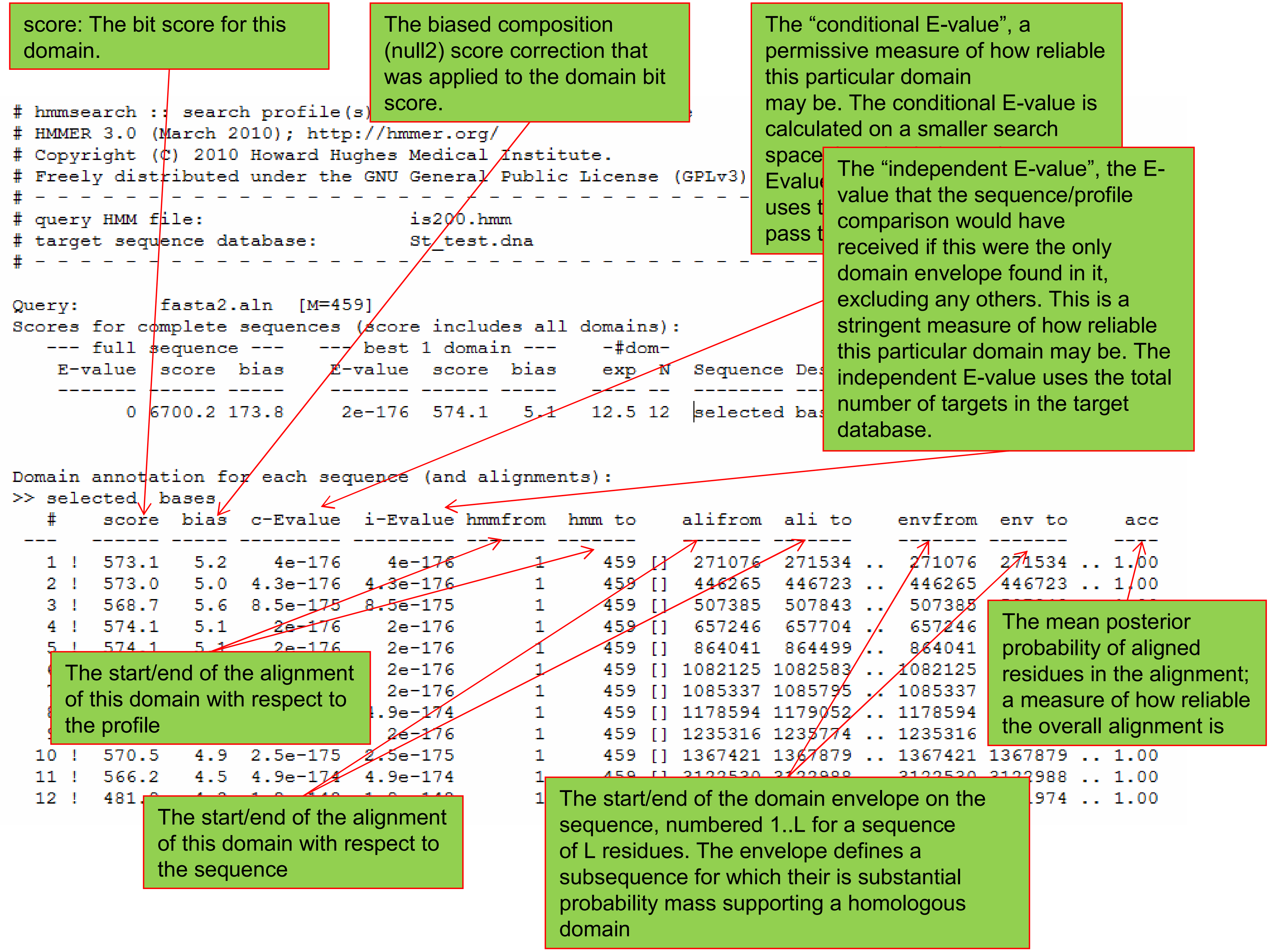


```
# hmmsearch :: search profile(s)
# HMMER 3.0 (March 2010); http://hmmer.org/
# Copyright (C) 2010 Howard Hughes Medical Institute.
# Freely distributed under the GNU General Public License (GPLv3)
# - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - -
# query HMM file:                  is200.hmm
# target sequence database:        St_test.dna
# - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - -

Query:       fasta2.aln  [M=459]
Scores for complete sequences (score includes all domains):
   --- full sequence ---   --- best 1 domain ---    -#dom-
    E-value  score  bias    E-value  score  bias    exp  N  Sequence Des
    ------- ------ -----    ------- ------ -----   ---- --  -------- ---
          0 6700.2 173.8     2e-176  574.1   5.1   12.5 12  selected bas


Domain annotation for each sequence (and alignments):
>> selected  bases
   #    score  bias  c-Evalue  i-Evalue hmmfrom  hmm to    alifrom  ali to    envfrom  env to     acc
 ---   ------ ----- --------- --------- ------- -------    ------- -------    ------- -------    ----
   1 !  573.1   5.2    4e-176    4e-176       1     459 []  271076  271534 ..  271076  271534 .. 1.00
   2 !  573.0   5.0  4.3e-176  4.3e-176       1     459 []  446265  446723 ..  446265  446723 .. 1.00
   3 !  568.7   5.6  8.5e-175  8.5e-175       1     459 []  507385  507843 ..  507385
   4 !  574.1   5.1    2e-176    2e-176       1     459 []  657246  657704 ..  657246
   5 !  574.1   5.1    2e-176    2e-176       1     459 []  864041  864499 ..  864041
   6 !  574.1   5.1    2e-176    2e-176       1     459 [] 1082125 1082583 .. 1082125
   7 !  574.1   5.1    2e-176    2e-176       1     459 [] 1085337 1085795 .. 1085337
   8 !  566.2   4.5  4.9e-174  4.9e-174       1     459 [] 1178594 1179052 .. 1178594
   9 !  574.1   5.1    2e-176    2e-176       1     459 [] 1235316 1235774 .. 1235316
  10 !  570.5   4.9  2.5e-175  2.5e-175       1     459 [] 1367421 1367879 .. 1367421 1367879 .. 1.00
  11 !  566.2   4.5  4.9e-174  4.9e-174       1     459 [] 3122530 3122988 .. 3122530 3122988 .. 1.00
  12 !  481.                                 1                                        974 .. 1.00
```

The mean posterior probability of aligned residues in the alignment; a measure of how reliable the overall alignment is

The start/end of the alignment of this domain with respect to the sequence

The start/end of the domain envelope on the sequence, numbered 1..L for a sequence of L residues. The envelope defines a subsequence for which their is substantial probability mass supporting a homologous domain



score: The bit score for this domain.

The biased composition (null2) score correction that was applied to the domain bit score.

The “conditional E-value”, a permissive measure of how reliable this particular domain may be. The conditional E-value is calculated on a smaller search space
Evalu
uses t
pass t

The “independent E-value”, the E-value that the sequence/profile comparison would have received if this were the only domain envelope found in it, excluding any others. This is a stringent measure of how reliable this particular domain may be. The independent E-value uses the total number of targets in the target database.

```
# hmmsearch :: search profile(s)
# HMMER 3.0 (March 2010); http://hmmer.org/
# Copyright (C) 2010 Howard Hughes Medical Institute.
# Freely distributed under the GNU General Public License (GPLv3)
# - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - -
# query HMM file:                  is200.hmm
# target sequence database:        St_test.dna
# - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - -

Query:       fasta2.aln  [M=459]
Scores for complete sequences (score includes all domains):
   --- full sequence ---   --- best 1 domain ---    -#dom-
    E-value  score  bias    E-value  score  bias    exp  N  Sequence Des
    ------- ------ -----    ------- ------ -----   ---- --  -------- ---
          0 6700.2 173.8     2e-176  574.1   5.1   12.5 12  selected bas


Domain annotation for each sequence (and alignments):
>> selected  bases
   #    score  bias  c-Evalue  i-Evalue hmmfrom  hmm to    alifrom  ali to    envfrom  env to     acc
 ---   ------ ----- --------- --------- ------- -------    ------- -------    ------- -------    ----
   1 !  573.1   5.2    4e-176    4e-176       1     459 []  271076  271534 ..  271076  271534 .. 1.00
   2 !  573.0   5.0  4.3e-176  4.3e-176       1     459 []  446265  446723 ..  446265  446723 .. 1.00
   3 !  568.7   5.6  8.5e-175  8.5e-175       1     459 []  507385  507843 ..  507385
   4 !  574.1   5.1    2e-176    2e-176       1     459 []  657246  657704 ..  657246
   5 !  574.1   5.1    2e-176    2e-176       1     459 []  864041  864499 ..  864041
                               2e-176       1     459 [] 1082125 1082583 .. 1082125
                               2e-176       1     459 [] 1085337 1085795 .. 1085337
                              .9e-174       1     459 [] 1178594 1179052 .. 1178594
                               2e-176       1     459 [] 1235316 1235774 .. 1235316
  10 !  570.5   4.9  2.5e-175  2.5e-175       1     459 [] 1367421 1367879 .. 1367421 1367879 .. 1.00
  11 !  566.2   4.5  4.9e-174  4.9e-174       1     459 [] 3122530 3122988 .. 3122530 3122988 .. 1.00
  12 !  481.                                  1                                          974 .. 1.00
```

The mean posterior probability of aligned residues in the alignment; a measure of how reliable the overall alignment is

The start/end of the alignment of this domain with respect to the profile

The start/end of the alignment of this domain with respect to the sequence

The start/end of the domain envelope on the sequence, numbered 1..L for a sequence of L residues. The envelope defines a subsequence for which their is substantial probability mass supporting a homologous domain