



Formatting your Submission

Created: April 6, 2011; Updated: November 3, 2014.

Sequence Formatting

FASTA Formatting

What is FASTA formatting

Define the FASTA format.

Simply speaking, the FASTA format consists of a single-line description of the sequence (called the definition line), which is followed by raw sequence data.

FASTA Formatting For Nucleotide Sequences

How do I format my nucleotide sequence in FASTA?

1. Open a text editor:
 - a. Microsoft (MS) operating system:
Open Notepad (Wordpad) or other MS compatible text editor.
 - Save document as plain text (.txt).
 - Do not save as rich text (.rtf) or as a document (.doc)
 - b. Mac operating system:
Open Textedit, TextWrangler or other Mac compatible text editor.
 - Save document as plain text (.txt).
 - Do not save as rich text (.rtf) or as a document (.doc).
2. Create the definition line for your sequence.
3. Press the “Enter” key to begin a new line following the definition line.
4. Enter your raw sequence data:
 - a. Total sequence size must be at least 200 bp (shorter sequences will not be processed).
 - b. Present your sequence data using Making Tab-delimited Tables symbols. Any non-IUPAC symbols (including dashes) will be removed from the sequence when it is imported into Sequin. BankIt will not accept sequences with non-IUPAC symbols
 - c. Letter case is ignored, so you may enter the sequence symbols in either upper or lower case. Any significance you attach to the case of the symbols will be lost.
 - d. Use the IUPAC approved "N" symbol to represent ambiguous sequence data.
 - e. Lines of sequence data in FASTA format should be 80 characters or shorter.

How do I format a FASTA definition line for a nucleotide sequence?

The FASTA definition line must be constructed in the following order and in a single line of text. Do not insert any hard returns (the “Enter” key on your keyboard) in the definition line. In both Sequin and BankIt,

the organism and any additional modifiers can be added as tables later in the submission process if you do not include them in the FASTA definition line. The SeqID is the only mandatory component of the FASTA definition line.

1. Type a carat (">") symbol.
2. Enter the sequence identifier (SeqID). This SeqID:
 - Must be unique for each nucleotide sequence
 - Cannot contain any spaces
 - Cannot contain brackets
 - Should be relatively short (preferably under 25 characters). Do not use the complete organism name for the SeqID.
 - Isolate, strain, clone or other laboratory identifiers are examples of SeqIDs.
3. Type a space.
4. Type information about the organism where you obtained the sequence:
 - This information must be in the format [modifier=text]. For example:

[organism=Gallus gallus]
 - Do not put spaces around the "="
5. At this point you can add additional information to describe the sequence in the form of optional **modifiers**:
 - a. Type a space following the information you entered in step 4.
 - b. Enter optional modifiers to describe the sequence:
 - This information must be in the format [modifier=text]. For example:
[breed=booted bantam]
 - Do not put spaces around the "="
6. At this point, you can add an optional title for your sequence:
 - a. Type a space.
 - b. Enter an optional descriptive title for your sequence:
 - Here is an example of a descriptive title for a sequence:
Gallus gallus doublesex and mab-3 related transcription factor 1 (DMRT1)
 - As GenBank has a preferred format for nucleotide and protein titles, the sequence title you provide will be changed to the proper format by the database staff during processing.
7. End your definition line by pressing the "Enter" key on your keyboard to insert a hard return.
8. Here is an example of a completed FASTA nucleotide sequence definition line whose components are the examples used in the steps above:
>SEQ1 [organism=Gallus gallus] [breed=booted bantam] doublesex and mab-3 related transcription factor 1 (DMRT1)
9. Begin entering your raw sequence data in the accepted FASTA format

Here is an example of how the sequence definition line will look followed by nucleotide sequence:

```
>SEQ1 [organism=Gallus gallus] [breed=booted bantam] doublesex and mab-3 related
transcription factor 1 (DMRT1)
CCGGCGGCGGGCAAGAAGCTGCCGCTGCCCCAAGTGTGCCCGCTGCCGCAACCACGGCTACTCCTCGC
CGCTGAAGGGGCACAAGCGGTTCTGCATGTGGCGGGACTGCCAGTGCAAGAAGTGCAGCCTGATCGCCGA...
```

Remember:

- Do not include any hard returns in your FASTA definition line (by hitting the "Enter" button on your keyboard) until the end of the definition line, or you may have trouble importing your FASTA sequences to GenBank.

- If you do have trouble importing your sequences, please double check that no returns were added to the FASTA definition line by your editing software.

What kind of information should I include in the definition line of my submission?

Although you can include any information you want in your definition line, the information you do include will be edited during GenBank processing to conform to specific database criteria, and therefore the definition line you provide will probably not remain the same after processing.

See the Sequin help documentation [section](http://www.ncbi.nlm.nih.gov/Sequin/QuickGuide/sequin.htm) <http://www.ncbi.nlm.nih.gov/Sequin/QuickGuide/sequin.htm> - [NucleotidePage](#) on importing nucleotide FASTA for more specifics on formatting a FASTA sequence importation file.

FASTA Formatting For Protein Sequences

How do I format my protein sequence in FASTA?

1. Open a text editor:
 - a. Microsoft (MS) operating system:
Open Notepad (Wordpad) or other MS compatible text editor.
 - Save document as plain text (.txt).
 - Do not save as rich text (.rtf) or as a document (.doc).
 - b. Mac operating system:
Open Textedit, TextWrangler or other Mac compatible text editor.
 - Save document as plain text (.txt).
 - Do not save as rich text (.rtf) or as a document (.doc).
2. Create the definition line for your sequence.
3. Press the “Enter” key to begin a new line following the definition line.
4. Enter your raw sequence data:
 - a. Present your sequence data using IUPAC symbols.
 - b. Letter case is ignored, so you may enter the sequence symbols in either upper or lower case. Any significance you attach to the case of the symbols will be lost.
 - c. Lines of sequence data in FASTA format should be 80 characters or shorter.

How do I format a FASTA definition line for a protein sequence?

Note: You will not need to import a protein sequence into BankIt or Sequin if the nucleotide spans for the coding region are provided.

The FASTA definition line must be constructed in the following order and in a single line of text. Do not insert any hard returns (the “Enter” key on your keyboard) in the definition line.

1. Type a carat (“>”)symbol, then a space.
2. Enter the sequence identifier (SeqID). This SeqID:
 - must be the same SeqID that you used to identify the nucleotide sequence.
 - In the case of alternatively spliced genes, a single protein FASTA file can contain two unique sequences that have the same SeqID. Both coding regions will be added to the same nucleotide sequence.
 - Cannot contain any spaces
 - Cannot contain brackets
3. Type a space.
4. Enter the protein name in the format [modifier=text]. For example:
[protein=doublesex and mab-3 related transcription factor 1]
 - Do not put spaces around the “=”

5. At this point you can add additional information to describe the sequence in the form of optional modifiers:
 - a. Type a space following the information you entered in step 4.
 - b. Enter optional modifiers to describe the sequence:
 The modifiers available for use in a protein FASTA definition line are different than those for a nucleotide FASTA definition line and are limited to the following information about the protein or gene itself, and should be presented in the format [modifier=text]:
 - [gene=text] Example:
[gene=DMRT1]
 - Do not put spaces around the “=”
6. End your definition line by pressing the “Enter” key on your keyboard to insert a hard return.
 - Here is an example of a completed FASTA protein sequence definition line whose components are examples used in the steps above:
 >SEQ1 [gene=DMTR1] [protein=doublesex and mab-3 related transcription factor 1]
7. Begin entering your raw sequence data in the accepted FASTA format.
 Here is an example of how the sequence definition line will look followed by protein sequence:
 >SEQ1 [gene=DMTR1] [protein=doublesex and mab-3 related
 transcription factor 1]
 PAAGKKLPRLPKCARCRNHGYSSPLKGHKRFCMWRDCQCKKCSLIAERQVRMAVQVALRRQQAQEEEL
 GI
 SHPVPLPSAPEPVVKSSSSSSCLLQDSSSPAHSTSTVAAAAASAPPEGRMLIQDIPSI PSRGHLEST
 SD...

Remember:

- Do not include any hard returns in your FASTA definition line (by hitting the “Enter” button on your keyboard) until you reach the end of your definition line, or you may have trouble importing your FASTA sequences to GenBank.
- If you do have trouble importing your sequences, please double check that no returns were added to the FASTA definition line by your editing software.

Formatting Sequence Gaps

How does GenBank define a sequence “gap”?

GenBank defines a sequence gap as

- A region of unknown sequence.

OR

- A region of un-sequenceable sequence that lies between two known regions of sequence.

If I don’t know the base at a particular position in my sequence data, can I use “-” or “?” to represent the unknown base?

You may use the - or ? characters in sequence data for alignment submissions only.

These symbols will be stripped from your sequence by our submission processing software if you include them in a FASTA file, so **you’ll need to insert a series of nnnns where each gap (see the answer to “How does GenBank define a sequence ‘gap’?) is located. If the gap length is estimated, insert the equivalent number of nns to represent the gap. If the gap length is unknown, insert 100 n’s.**

Note: GenBank cannot accept sequences where 50% or more of the submitted sequence is represented by internal Ns (See the answer to “Can I submit a sequence to GenBank that has gaps in it?” for more information about formatting internal Ns).

Below is a [IUPAC](#) (International Union of Pure and Applied Chemistry) code table for your reference.

IUPAC-IUB single-letter base codes:

Code	Base Description
G	Guanine
A	Adenine
T	Thymine (Uracil in RNA)
C	Cytosine
R	Purine (A or G)
Y	Pyrimidine (C or T or U)
M	Amino (A or C)
K	Ketone (G or T)
S	Strong interaction (C or G)
W	Weak interaction (A or T)
H	Not-G (A or C or T) H follows G in the alphabet
B	Not-A (C or G or T) B follows A in the alphabet
V	Not-T (not-U) (A or C or G) V follows U in the alphabet
D	Not-C (A or G or T) D follows C in the alphabet
N	Any (A or C or G or T)

[Pure and Applied Chemistry 40 \(3\) , 277 - 331 \(1974\)](#)

[Nomenclature Committee of the International Union of Biochemistry. Ref: Cornish-Bowden, A. Nucl Acid Res 13, 3021-3030 \(1985\)](#)

Can I submit a sequence to GenBank that has gaps in it? If so, how do I represent the gaps?

For sequences that are from the same organism and individual, and are part of the same gene or locus, but have some sequence missing (like exons of a gene, where the introns are missing), **you'll need to insert a series of nnnns where each gap (see the answer to “How does GenBank define a sequence ‘gap?’) is located.**

- **If the gap length is estimated**, insert the equivalent number of nnnns to represent the gap
- **If the gap length is unknown**, insert a string of 100 nnnns to represent the gap
- **Annotate each gap as a misc_feature** (miscellaneous feature) and include a note describing each gap. For gaps of unknown length, be sure to include in your note an explanation that describes the region(s) or feature(s) that is missing (i.e. the missing sequence represented by the nnnns in your gapped submission sequence). **For example:**
 - /note="gap, unknown length", intron 2"
 - /note="gap, estimated length, ## base pairs"

Use the [gap specifications](#) provided in the Sequin Help documentation when you set-up your FASTA-formatted file for importation into Sequin.

BankIt users follow the bulleted points listed above.

Note: GenBank cannot accept sequences where more than 50% of the submitted sequence is gapped (represented by internal nnns).

Sequence Size

Does GenBank have a minimum size requirement for submitted sequences?

GenBank will process nucleotide sequences submissions that are $\geq 200\text{bp}$ in length (Sequences $< 200\text{bp}$ are accepted if they represent complete small RNAs or exons).

IUPAC Use

What are the IUPAC codes for nucleotides?

Code	Base Description
G	Guanine
A	Adenine
T	Thymine (Uracil in RNA)
C	Cytosine
R	Purine (A or G)
Y	Pyrimidine (C or T or U)
M	Amino (A or C)
K	Ketone (G or T)
S	Strong interaction (C or G)
W	Weak interaction (A or T)
H	Not-G (A or C or T) H follows G in the alphabet
B	Not-A (C or G or T) B follows A in the alphabet
V	Not-T (not-U) (A or C or G) V follows U in the alphabet
D	Not-C (A or G or T) D follows C in the alphabet
N	Any (A or C or G or T)

Pure and Applied Chemistry **40** (3) , 277 - 331 (1974)

Nomenclature Committee of the International Union of Biochemistry. Ref: Cornish-Bowden, A. Nucl Acid Res 13, 3021-3030 (1985).

Any IUPAC (International Union of Pure and Applied Chemistry) approved single-letter base code for nucleotides, including N, is acceptable for nucleotide sequence data submitted to GenBank.

What are the IUPAC codes for amino acids?

Code	Amino Acid	Code	Amino Acid
A	alanine	P	proline
B	aspartate or asparagine	Q	glutamine
C	cystine	R	arginine
D	aspartate	S	serine
E	glutamate	T	threonine
F	phenylalanine	U	selenocysteine

Table continued from previous page.

Code	Amino Acid	Code	Amino Acid
G	glycine	V	valine
H	histidine	W	tryptophan
I	isoleucine	Y	tyrosine
K	lysine	Z	glutamate or glutamine
L	leucine	X	any amino acid
M	methionine		
N	asparagine		

IUPAC-IUB Joint Commission on Biochemical Nomenclature, Nomenclature and Symbolism for Amino Acids and Peptides [section 3AA-1](#): Names of common α -Amino Acids.

Any IUPAC (International Union of Pure and Applied Chemistry) approved single-letter base code for nucleotides, including X, is acceptable for nucleotide sequence data submitted to GenBank.

Making Tab-delimited Tables

What is a tab-delimited table?

A **tab-delimited table** is a table where a single tab keystroke “delimits” (marks the boundary) between one column and the next in a table.

The format requirements of each tab-delimited table are different, and therefore you should consult the specific table instructions for the resource you are using before you begin your table. Regardless of the type of tab-delimited table you are making, **always follow these rules when making a tab-delimited table**:

- Do **not** use more than one tab keystroke between columns in the table to make the data in the columns align.
- Do **not** use the space bar (in addition to your single tab keystroke) between columns in the table to make the data in the columns align.
- When you save the table:
- **Save the table as plain text (.txt).**
- Do not save the table as rich text (.rtf) or as a document (.doc)

See the **answer to** “Can you give me step-by-step instructions for making a tab-delimited feature table...”, located in this section to see step-by-step **instructions for making a tab-delimited feature table**.

Feature Table

Can you give me step-by-step instructions for making a tab-delimited feature table for my GenBank submission?

A **tab-delimited feature table** uses a single “Tab” keystroke to delimit (mark the boundary) between one column and the next in a table that contains your feature information.

Follow these instructions to make a tab-delimited feature table:

1. Open a text editor or spreadsheet program:
 - a. Microsoft (MS) operating system:
Open Notepad (Wordpad) or other MS compatible text editor
 - Save document as plain text (.txt).

- **Do not save** as rich text (.rtf) or as a document (.doc)
 - b. Mac operating system:
Open Textedit, TextWrangler or other Mac compatible text editor.
 - Save document as **plain text (.txt)**.
 - **Do not save** as rich text (.rtf) or as a document (.doc)
 - c. Spreadsheet program:
Open program, enter your data.
 - Save document as **plain text (.txt)**.
- 2. Create a Sequence ID (SeqID) Row.
The SeqID Row tells our submission system that a new set of features for the SeqID specified in this row will follow. The SeqID Row contains the following:
 - >Features, a space (hit the spacebar on your keyboard once), and the SeqID of the sequence you are annotating. In the example below, eIF4E is the SeqID used in the FASTA file for the sequence:

```
>Features lcl|eIF4E
```
- 3. Hit the Enter key of your keyboard once to go to the next row.
- 4. Create a Feature Row:
A Feature Row begins the column portion of the table. The table is composed of five columns (Start, Stop, Feature, Modifier and Modifier value), where each column is separated from the columns beside it by a **single tab keystroke (represented here by <tab>)**.
 - a. The Feature Row provides the span (start and stop values) and the type of **feature** you are supplying for the SeqID indicated in the SeqID Row:
 - SeqID Row: >Features SequenceID
Feature Row: **Start <tab> Stop <tab> Feature**
 - Here is an example of how the Feature Row would look in a text editor following the SeqID Row (See Box 1):
 - b. Additional intervals (if any) for a particular feature will appear in the rows following the Feature Row, where each interval contained in the feature is represented by its start value and the stop value (span) in its own row.

```
Feature Row: Start value <tab> Stop value <tab> Feature
Interval Row: Start value <tab> Stop value
Interval Row: Start value <tab> Stop value
Interval Row: Start value <tab> Stop value
```

Here is an example of how Feature Row and its additional intervals would look in a text editor (See Box 2).

- 5. Hit the Enter key of your keyboard once to go to the next row.
- 6. Create a Modifier Row:
A Modifier Row comes at the conclusion of a Feature Row (and any associated Interval Rows), and contains the modifier information for the Feature described in the row(s) above it. The Modifier Row provides the type of **modifier** as well as the value for that modifier. The Modifier Row begins with three tab keystrokes followed by the modifier name and then by the modifier value.

If you do not have modifiers to describe the feature provided in the Feature Row, skip down to step 8

- a. This is how you would enter a Modifier Row directly following a Feature Row for a particular SeqID:

SeqIDRow: >Features SequenceID
 Feature Row: **Start value** <tab> **Stop value** <tab> **Feature**
 Modifier Row: <tab> <tab> <tab> **Modifier** <tab> **Modifier value**

Here is an example of how the Modifier Row would look row in a text editor when it follows directly after the feature (See Box 3).

- b. This is how you would enter a Modifier Row following a Feature Row and its additional intervals:

Feature Row: **Start value** <tab> **Stop value** <tab> **Feature**
 Interval Row: **Start value** <tab> **Stop value**
 Interval Row: **Start value** <tab> **Stop value**
 Interval Row: **Start value** <tab> **Stop value**
 Interval Row: **Start value** <tab> **Stop value**
 Modifier Row: <tab> <tab> <tab> **Modifier** <tab> **Modifier value**

Here is an example of how the Modifier Row would look in a text editor when it follows a Feature Row and its Interval Rows (See Box 4).

7. Hit the Enter key of your keyboard once to go to the next row.
8. At this point you can do any one of the following:
 - Create another Modifier Row to provide more information for the feature you described in the Feature Row.
 - Create another Feature Row and proceed to describe the intervals and modifiers for this feature using Interval Rows (if any) and Modifier Rows (if any).
 - Create a new SeqID Row, and proceed to describe the features and modifiers for this SeqID using Feature Rows, Interval Rows (if any) and Modifier Rows (if any).
 - Here is how the examples used in the steps above would look together:

SeqIDRow: >**Features SequenceID**
 Feature Row: **Start value** <tab> **Stop value** <tab> **Feature**
 Modifier Row: <tab> <tab> <tab> **Modifier** <tab> **Modifier value**
 Feature Row: **Start value** <tab> **Stop value** <tab> **Feature**
 Interval Row: **Start value** <tab> **Stop value**
 Interval Row: **Start value** <tab> **Stop value**
 Interval Row: **Start value** <tab> **Stop value**
 Interval Row: **Start value** <tab> **Stop value**
 Modifier Row: <tab> <tab> <tab> **Modifier** <tab> **Modifier value**

Here is how the examples used in the steps above would look together in a text editor (See Box 5):

9. Once your table is imported into Sequin (or BankIt), Sequin/BankIt will recognize the SeqIDs in your table, and will automatically assign and place the appropriate features and their modifiers on each sequence in your set

When you create your feature table, **always remember the following:**

- When you make your tables:
 - Do **not** use more than one tab keystroke between columns in the table to make the data in the columns align.

- Do **not** use the space bar (in addition to your single tab keystroke) between columns in the table to make the data in the columns align.
- When you save the table:
 - Save the table as **plain text (.txt)**.
 - Do not save the table as rich text (.rtf) or as a document (.doc)

You can see the [complete feature table for lcl|eIF4E](#) (the example used above) in the Sequin Quick Guide, and you can also find an additional example of a more complex feature table in the “[Submission of Annotation using a Table](#)” page (scroll down to see the example in Fig. 1).

Box 1.

```
>Features lcl|eIF4E
80      2881   gene
```

Box 2.

```
201     224     CDS
1550    1920
1986    2085
2317    2404
```

Box 3.

```
>Features lcl|eIF4E
80      2881   gene
                                gene    eIF4E
```

Box 4.

```
201     224     CDS
1550    1920
1986    2085
2317    2404
2466    2629
                                product    eukaryotic initiation factor 4E-II
```

Box 5.

```
>Features lcl|eIF4E
80      2881   gene
                                gene    eIF4E

201     224     CDS
1550    1920
1986    2085
2317    2404
2466    2629
                                product    eukaryotic initiation factor 4E-II
```

Source Modifier Table

Can you give me step-by-step instructions for making a tab-delimited source modifier table for my GenBank submission?

Note: Due to a technical issue with the right margin that cannot be fixed, the example lines in this “step-by-step” for the source modifier table have been broken into two lines, but it is important that when you enter lines like these, they should be in a single line without breaks.

A tab-delimited source modifier table uses a single “Tab” keystroke to delimit (mark the boundary) between one column and the next in a table that contains your source modifier information.

Follow these instructions to make a tab-delimited source modifier table:

1. Open a text editor or spreadsheet program:
 - a. Microsoft (MS) operating system:
Open Notepad (Wordpad) or other MS compatible text editor.
 - Save document as plain **text (.txt)**.
 - Do not save as rich text (.rtf) or as a document (.doc).
 - b. Mac operating system:
Open Textedit, TextWrangler or other Mac compatible text editor.
 - Save document as plain **text (.txt)**.
 - Do not save as rich text (.rtf) or as a document (.doc).
 - c. Spreadsheet program:
Open program, enter your data.
 - Save document as **plain text (.txt)**.
2. **Create the Column Label Row:**
 - The First column always lists the Sequence IDs; each subsequent column in the table lists a different source modifier that will be applied. You will find a comprehensive [list of source modifiers](#) online.
 - a. Type this label: **Sequence_ID** as the first entry in the Column Label Row. The Sequence_ID must be the same as that used to identify each sequence in your nucleotide FASTA file.
 - b. Separate the Sequence_ID label by a single tab key stroke from the next column, which will be a source modifier label. Add as many source modifier labels to this row as you need, each separated from the next by a single tab keystroke. The column labels can be in whatever order you want so long as the Sequence_ID label starts the Column Label Row.
 - i. This is how you would enter a Column Label Row:

Column Label Row:
Sequence_ID<tab>Specimen_voucher<tab>Collected_by<tab>
Collection_date<tab>Country<tab>Identified_by <tab> Lat_Lon
 - ii. Here is an example of how the Column Label Row would look in a text editor (see Box 6)
3. **Hit the Enter key of your keyboard** once to go to the next row.
4. **Create a Source Modifier Row** for your first sequence:
 - a. Enter the sequence ID of your first sequence in the first column of the table. Enter a single tab keystroke, followed by the source modifier value for the source modifier column that follows the Sequence_ID column.
 - b. Enter another tab keystroke followed by the source modifier value for the second source modifier column that follows the Sequence_ID column. Continue to enter the value data for

each source modifier label (each value separated by a single tab keystroke from the next) until all the source modifier data for the first SeqID has been entered.

- i. This is how you would enter a Source Modifier Row following a Column Label Row:

Column Label Row: Sequence_ID<tab>Specimen_voucher<tab>Collected_by<tab>Collection_date<tab>Country<tab>Identified_by <tab>Lat_Lon

Source Modifier Row 1: Seq1<tab>MKP 334<tab>C. Grant<tab>31-Jan-2001 USA<tab> C. Grant<tab>13.57N 24.68 W

- ii. Here is an example of how the Column Label Row and Source Modifier Row would look in a text editor (see Box 7).

5. Create a Source Modifier Row for your second sequence:

- a. Enter the sequence ID of your second sequence in the first column of the table. Enter a single tab keystroke, followed by the source modifier value for the source modifier column that follows the Sequence_ID column.

- b. Enter another tab keystroke followed by the source modifier value for the second column that follows the Sequence_ID column. Continue to enter data for each source modifier label (each separated by a single tab keystroke from the next) until all the source modifier data for the second SeqID has been entered.

- i. This is how you would enter a second Source Modifier Row following a Column Label Row and a preceding Source Modifier Row:

Column Label Row: Sequence_ID<tab>Specimen_voucher<tab>Collected_by<tab>Collection_date<tab>Country<tab>Identified_by <tab>Lat_Lon

Source Modifier Row 1: Seq1<tab> MKP 334<tab>C. Grant<tab>31-Jan-2001 USA<tab> C. Grant<tab> 13.57N 24.68 W

Source Modifier Row2: Seq2<tab>MKP 1230<tab>S. Tracy<tab> 28-Feb-2002<tab>Slovakia <tab>C. Grant<tab>13.24 N 24.35 W

- ii. Here is an example of how the Column Label Row and Source Modifier Rows would look in a text editor (See Box 8).

- Because you are using Tabs to separate columns, each Source Modifier Row column may not line up with the Column Label Row columns or with other Source Modifier Row columns. This is normal and valid, as long as you are using only Tabs between each column.
- Do **not** use more than one tab keystroke between columns in the table to make the data in the columns align.
- Do **not** use the space bar (in addition to your single tab keystroke) between columns in the table to make the data in the columns align.

6. Continue to add Source Modifier Rows for each of your remaining sequences.

- a. This is how you would enter an additional Source Modifier Rows following a Column Label Row and preceding Source Modifier Rows:

Column Label Row: Sequence_ID<tab>Specimen_voucher <tab>Collected_by<tab>Collection_date<tab>Country<tab> Identified_by<tab>Lat_Lon

Source Modifier Row 1: Seq1<tab>MKP 334 <tab>C. Grant<tab>31-Jan-2001<tab>USA<tab> C. Grant<tab>13.57 N 24.68 W

Source Modifier Row 2: Seq2<tab>MKP 1230 <tab>S. Tracy<tab>28-Feb-2002<tab>Slovakia <tab>C. Grant<tab>13.24 N 24.35 W

Source Modifier Row 3: Seq3<tab>1B-2526<tab>A. Gardner<tab> 16-Apr-2001France<tab>C. Grant <tab>43.21 N 56.78 W

Source Modifier Row 4: Seq4<tab>WBM 86-64<tab>F. McMurray<tab> 26-May-2002<tab>Germany<tab>C. Grant<tab>45.32 N 21.34 E

Source Modifier Row 5: Seq5<tab> 1B-2518<tab>V. Leigh<tab>13-Jun- 2003<tab>Brazil<tab>V. Leigh<tab>46.80 N 13.57 E

- b. Here is an example of how the Column Label Row and additional Source Modifier Rows would look in a text editor (See Box 9.)
- 7. Once your table is imported into Sequin (or BankIt), Sequin/BankIt will recognize the SeqIDs in your table, and will automatically assign and place the appropriate source modifiers on each sequence in your set.
- 8. When you create your Source Modifier table, **always remember the following:**
 - Do **not** use more than one tab keystroke between columns in the table to make the data in the columns align.
 - Do **not** use the space bar (in addition to your single tab keystroke) between columns in the table to make the data in the columns align.
 - Each Sequence ID (SeqID) can appear **only once** in a source modifier table.
 - When you save the table:
 - Save the table as **plain text** (.txt).
 - Do **not** save the table as rich text (.rtf) or as a document (.doc).

Please see the BankIt Submission [Help documentation](#) for further information on creating source modifier tables.

Box 6.

```
Sequence_ID      Specimen_voucher      Collected_by      Collection_date      Country
Identified_by    Lat_Lon
```

Box 7.

```
Sequence_ID      Specimen_voucher      Collected_by      Collection_date      Country
Identified_by    Lat_Lon
Seq1             MKP 334               C. Grant           31-Jan-2001         USA                C. Grant           13.57
N 24.68 W
```

Box 8.

```
Sequence_ID      Specimen_voucher      Collected_by      Collection_date      Country
Identified_by    Lat_Lon
Seq1             MKP 334               C. Grant           31-Jan-2001         USA                C. Grant           13.57 N 24.68 W
Seq2             MKP 1230              S. Tracy           28-Feb-2002         Slovakia           C. Grant           13.24 N
24.35 W
```

Box 9.

```
Sequence_ID      Specimen_voucher      Collected_by      Collection_date      Country
Identified_by    Lat_Lon
Seq1             MKP 334               C. Grant           31-Jan-2001         USA                C. Grant           13.57 N
24.68 W
Seq2             MKP 1230              S. Tracy           28-Feb-2002         Slovakia           C. Grant           13.24 N
24.35 W
```

Box 9 continues on next page...

Box 9 continued from previous page.

	Seq3	1B-2526	A. Gardner	16-Apr-2001	France	C. Grant	43.21
N 56.78	W						
	Seq4	WBM 86-64	F. McMurray	26-May-2002	Germany	C. Grant	
45.32	N 21.34	E					
	Seq5	1B-2518	V. Leigh	13-Jun-2003	Brazil	V. Leigh	46.80
13.57	E						N