

Applied Biosystems Genetic Analysis Data File Format

July 2006

SUBJECT: ABIF File Format Specification and Sample File Schema

In This Document This document includes the following topics:

Introduction	2
A Note About Support	2
Important Notes About Compatibility	3
The ABIF File Format	4
Detailed Structure of the ABIF File	7
Sample File Schemas	22
ABI PRISM® 3100 and 3100-Avant™ Genetic Analyzer Tags	23
Applied Biosystems 3130/3130xl Genetic Analyzer Tags	29
Applied Biosystems 3730/3730xl DNA Analyzer Tags	36
SeqScape® Software v2.5 Tags	50
Sequencing Analysis Software v5.2 Tags	52

Introduction

This document is intended for programmers or bioinformatics groups who wish to perform additional analysis or other manipulation of ab1 and/or fsa files. The ab1 file is a file type produced by Data Collection software generating sequencing data, with the extension “.ab1”. The fsa file is a file type produced by Data Collection software generating fragment analysis data, with the extension “.fsa”. Both the ab1 and fsa files use the ABIF file format.

The ABIF file format specifies the general rules on how the file is constructed, and therefore the rules on how it can be read. Elements of data stored in the file are associated with tags, which are analogous to the keys in a (key, value) mapping.

The ABIF file format by itself does not specify the schema for the ab1 and fsa files, i.e. which tags are written and when. These schema are specific to the instrument and software version which created the file.

This document describes the ABIF format. Following the ABIF specification are the schemas for each instrument-software combination, for both the ab1 and fsa files starting on page 23.

Schemas (tables with the valid tags) for the following instruments are given:

- ABI PRISM® 3100 and 3100-Avant™ Genetic Analyzer Tags (page 23)
- Applied Biosystems 3130/3130xl Genetic Analyzer Tags (page 29)
- Applied Biosystems 3730/3730xl DNA Analyzer Tags (page 36)

A Note About Support

IMPORTANT! Applied Biosystems does *not* support users of this specification in any way. Please do not call technical support for additional information pertaining to this specification.

Important Notes About Compatibility

Backward and Cross Compatibility Some tags exist in the ab1 and fsa files for backward compatibility with earlier versions of Applied Biosystems software. They are no longer used by current versions of downstream analysis applications.

The ab1 and fsa schema documentation is provided for a specific instrument model and software version. There is no guarantee that the tagged data described will be consistent with files produced by earlier software releases and/or other instrument models.

Forward Compatibility The critical data in the ab1 and fsa files is stable, and in general new data will extend the existing schema. However, Applied Biosystems provides no guarantee that all tagged data elements will be present, consistent, or supported in future versions of the software, particularly for data that pertains to the details of instrument control or software integration.

Compatibility of Edited Sample Files There are two ways to modify sample files (ab1 and fsa files), either by adding new tags or changing existing tags. Sample files with new tags added by a user following Applied Biosystems' instructions as set forth in "Detailed Structure of the ABIF File" on page 7, should continue to be compatible with Applied Biosystems software. Any modification to sample files by changing the existing tags may result in the file no longer being compatible with Applied Biosystems software.

IMPORTANT! Applied Biosystems *does not* recommend any modification of the software files. Applied Biosystems *does not* support the editing of sample files in any way and *makes no guarantees* as to the compatibility of such files with Applied Biosystems software.

The ABIF File Format

Introduction The ABIF file format is a binary file format for storing data. Elements of data stored in the file are associated with tags, which are analogous to the keys in a (key, value) mapping.

The ABIF format can accommodate a moderate number (<1000) of heterogeneous data items. A data item can be a scalar value or an array. The basic data types are 8, 16, and 32 bit integers, 32 bit floating-point values, and ASCII characters. There are also two compound data types: date and time. The data type of each item is identified by an element type code.

Each data item is uniquely identified (tagged) within a file by a tag name and a tag number. The tag name and tag number are stored internally as 32-bit integers; the tag name is intended to be defined as a string of four 8-bit ASCII characters stored in big-endian order. For example, the tag named ABCD is represented by the hex value 0x41424344. The ABIF format also includes a directory of all the tagged data elements that are contained in a particular file.

Compatibility Goals Previous versions of ABIF libraries implicitly defined a complex format with many features that may have been under utilized.

The goal of this document is to reduce this complexity by creating a restricted definition that still meets the following goals:

- Provide full read and write support for all data types needed by current (2005) applications.
- Provide full read and write support for the thumbprint and boolean legacy data types.
- Provide read and write support for user types only in the form of raw byte arrays.
- Provide read support only for all other legacy data types, only in the form of raw byte arrays.
- Ensure that any file written according to this specification can be read by the current (2005) Applied Biosystems software.
- Ensure that any ABIF file written by an application released during or after 1998 can be read according to this specification.
- Eliminate any ambiguity in implementation requirements.

Historical Notes Applied Biosystems modeled the ABIF format after Tag Image File Format (TIFF), a format for graphics files, and the Macintosh OS Resource Manager. “ABIF” is an abbreviation of Applied Biosystems, Inc. Format.

The original ABIF specification was written in an era when a typical computer had 1 MB of RAM and operated at 16 MHz. Therefore, the ABIF libraries were designed to perform input/output operations in several small pieces to minimize the amount of data resident in RAM at one time.

The early ABIF file was expected to also serve as a simple database or nonresident data structure, since virtual memory was not a feature of the operating system at that time. The format was originally implemented on the Classic Mac OS, which used floating blocks of memory called handles. This is the origin of the datahandle field in ABIF directory entries. The datahandle field is reserved for internal use by libraries; it has no meaning in the file itself, but it should not be modified or used for any other purpose. Part of the header area was reserved for managing range-locking of data items. This was part of a plan to implement multi-user access controls, which were never implemented.

To avoid the effort involved in rewriting a file from scratch, the original ABIF specification allowed for multiple, linked tag directories (as does TIFF) so that data could easily be appended to an existing file. Also like TIFF, ABIF originally allowed for little-endian (as well as big-endian) byte ordering, which would be indicated by the order of the letters (A B I F) in the first four bytes of the file. These features were probably never used, and they have been eliminated from the current specification.

Overview of the ABIF Format

Data Types The ABIF format supports storage of the basic low-level data types common to most programming languages. These include char (a single byte character), short (a two byte integer), long (a four byte integer), float (a four byte floating point value) and double (an eight byte floating point value). Data stored as any of these basic types can be either scalar values or arrays.

In addition, two types of strings are supported. The cString type is a C-style string (null terminated). The pString type is a Pascal-style string (the length of the string is stored in the first byte of data). Values stored as pString are required to be less than 256 characters.

Some additional storage formats have been defined and are described below.

Data Tags

Tags are used to index the data contained in the file and can be thought of as (name, number) pairs. In practice, the names are required to be four characters and thus can always be converted to four-byte integers.

Unique (name, number) combinations define unique tags. For example, tags with the same name but different numbers are allowed to be of different types and contain unrelated information.

Data Storage

A designated section of the ABIF file contains a directory. The directory entries contain the tag (name, number) information, data type, number of elements, etc. For data values that are four bytes or less, the value is stored in the directory entry. Otherwise, an offset to the binary data in the file is stored. Details of the directory and binary storage formats are described below.

Detailed Structure of the ABIF File

The file consists of three logical parts:

- Header – discussed below
- Directory – discussed on page 10
- Data – discussed on page 13

The header area is located at a fixed position at the beginning of the file. The data and directory areas have no fixed locations.

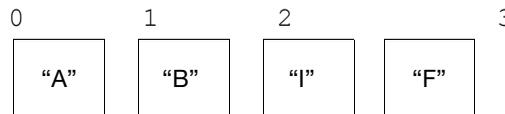
Although it may be convenient to do so, there is no requirement that the directory be located at the end of the file. The directory may be located before, after, or in between other data items, and there may be unused space between items.

All integers in ABIF data structures are stored in big-endian order (high-order byte first).

In the following descriptions, `SInt16` refers to a signed 16-bit integer, `UInt8` refers to an unsigned 8-bit integer, etc.

Header

File Signature The header is a 128-byte area at the beginning of the file.



The first four bytes of the file are the ASCII codes for A, B, I, F. Your implementations should check these bytes to verify that a file's format is ABIF.

Version Number 4



The next two bytes comprise a 16-bit integer corresponding to the version number of the format.

The version number is listed in earlier libraries as being equal to “version number field x100”, suggesting that the current value of 101 would be interpreted as “version 1.01,” i.e., the first minor variation number of the first major version.

A common interpretation of major and minor version numbers is that a major version change indicates a break in code compatibility, while a minor version change indicates a change only in interpretation or content. Given that, files conforming to the specification in this document would have a version number of 102, because the major compatibility is the same, but some obsolete features have been formally dropped. This is not the case, however. As long as compatibility with Applied Biosystems software is required, your implementations must continue to write a value of 101 here.

Your implementation must read this value to check for compatibility between the file’s format and the current version of the library; perform the check by dividing this value by 100 to get the major version number, and then comparing that value with the major version of the library. If the values differ, your implementation must return an error without attempting to read further. If the major version numbers are the same, reading may continue and it is up to the client application to handle any difference in minor version number.

Directory Entry Structure

The next 28 bytes comprise a single directory entry structure that points to the directory. A directory entry is a packed structure (no padding bytes) of the following form:

```
struct DirEntry{
    SInt32 name;          //tag name
    SInt32 number;       //tag number
    SInt16 elementtype; //element type code
    SInt16 elementsize; //size in bytes of one
                        element
    SInt32 numelements; //number of elements in item
    SInt32 datasize;    //size in bytes of item
    SInt32 dataoffset;  //item’s data, or offset
                        in file
    SInt32 datahandle;  //reserved
```

}

The DirEntry Structure is shown schematically below.

6

'tdir'

10

1

14

1023

16

28

18

num elements

22

data size

26

data offset

30

0

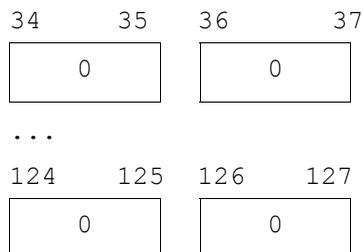
Your implementations which write ABIF should use the values shown above for tag number, element type and the other items in the DirEntry struct. The directory size (`datasize`) should be exactly the size required for the entries (`numelements` x `elementsiz`).

Your implementations that read ABIF must extract the `numelements` field, a 32-bit integer at byte 18, and the `dataoffset` field, a 32-bit integer at byte 26. These specify the number of entries in the directory and the location of the directory. The other fields should be ignored.

Note: Previous libraries may have reserved additional space in the directory, and therefore `datasize` may be larger than `numelements` x `elementsiz`.

Unused Space in the Header

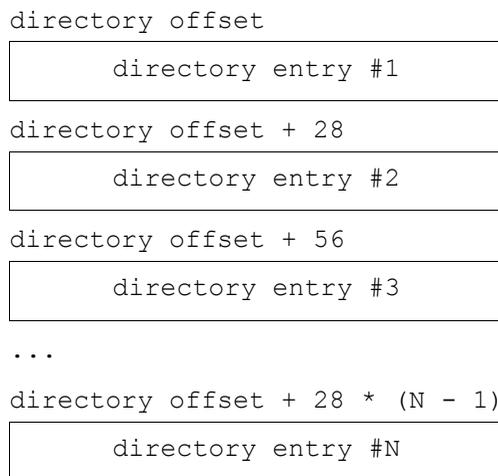
The DirEntry is followed by 47 2-byte integers, all to be ignored on input and set to zero on output.



The original spec reserved these fields to implement range-locking for a multi-user access scheme, but that feature was never implemented.

Directory

The directory is located at the offset specified in the header, and consists of an array of directory entries.



Fields in a Directory Entry

Each 28-byte entry consists of the following fields:

Name

```
SInt32 name; // tag name
```

The `name` field is defined as an integer but this field should be treated as an array of four 8-bit ASCII characters. Use printable ASCII characters in the range 0x21 to 0x7E, and it is appropriate to use an mnemonic abbreviation that is descriptive of the data item.

Tag Number

```
SInt32 number; // tag number
```

The `number` field can be any signed 32-bit integer, but it is customary to use positive values only, beginning with 1. It is also customary to use values less than 1000.

Element Type

```
SInt16 elementtype; // element type code
```

The `elementtype` indicates the type of data contained in the data item. New applications writing ABIF files should only use codes for current data types. (See “Current Data Types” on page 13.)

Your implementations should provide for writing user types, but only in the form of byte-arrays of raw data. The basic library should not provide interpretation or translation of unsupported types. (You have the option of defining custom user types in applications, but Applied Biosystems discourages this practice; see “Notes on Using the ‘User’ Data Type” on page 20.)

Your implementations should provide for reading all unsupported legacy element types (including user types ≥ 1024), but only in the form of byte-arrays of raw data. No interpretation or translation of unsupported types should be provided in the basic library. (These types are listed in “Unsupported Legacy Data Types” on page 17.)

Your implementations should also provide for editing or copying an existing file that *contains* tags with unsupported element types. Implementations should not provide for editing or creating new individual tags that use an unsupported element type.

If your implementation encounters an undefined element type (neither current nor legacy), it should report an error to the caller.

Library implementations should not introduce new data types or resurrect legacy data types.

Element Size

```
SInt16 elementsize; // size in bytes of one element
```

For all supported data types, the `elementsiz` field is redundant, since the element size for each type is uniquely defined by the specification. You may use or ignore this field on input in your implementations, but you must write the correct value on output.

For user types and unsupported types, ignore this value on input.

Implementations should neither make this field available to client software nor allow client software to set this value.

Number of Elements

```
SInt32 numelements; // number of elements in item
```

The `numelements` field gives the number of elements in the data item. Note that for the string types, an “element” is an individual character, not the string itself.

For user types and unsupported types, ignore this value on input and copy it to the file on output without use or interpretation.

Item's Size

```
SInt32 datasize; // size in bytes of item
```

The `datasize` field gives the number of bytes in the data item.

Offset to Item's Location

```
SInt32 dataoffset; // item's data, or offset in file
```

For data items of size greater than 4 bytes, the `dataoffset` field contains the offset to the data in the file.

For data items of 4 bytes or less, the `dataoffset` field contains the data item itself. In this case, the data bytes are stored beginning at the high-order byte of the 32-bit field. Examples:

Data Item	Dataoffset value
one byte with value	15 0x0F000000
two character pString "AB"	0x02414200
array of two shorts {1, 2}	0x00010002

Data Handle

```
SInt32 datahandle; // reserved
```

Your implementation should ignore the `datahandle` field on input and write a value of zero on output. (This field was used in libraries that implemented a virtual memory system similar to that of the Classic Mac OS Resource manager; see “Historical Notes” on page 5.)

Data Types

This specification describes three data types:

- Current types (see below)
- Legacy data types which should be supported (see page 16)
- Legacy data types which do not need to be supported (see page 17)

Current Data Types

Name	byte
Element type	1
Element size	1 byte
Description	Unsigned 8-bit integer.

Name	char
Element type	2
Element size	1 byte
Description	8-bit ASCII character or signed 8-bit integer.

Name	word
Element type	3
Element size	2 bytes
Description	Unsigned 16-bit integer.

Name	short
Element type	4
Element size	2 bytes
Description	Signed 16-bit integer.

Name	long
Element type	5
Element size	4 bytes
Description	Signed 32-bit integer.

Name	float
Element type	7
Element size	4 bytes
Description	32-bit floating point value.

Name	double
Element type	8
Element size	8 bytes
Description	64-bit floating point value.

Name	date
Element type	10
Element size	4 bytes
Description	Packed structure to represent calendar date: <pre> { SInt16 year; // 4-digit year UInt8 month; // month 1-12 UInt8 day; // day 1-31 } </pre>

Name	time
Element type	11
Element size	4 bytes
Description	Packed structure to represent time of day: <pre> { UInt8 hour; // hour 0-23 UInt8 minute; // minute 0-59 UInt8 second; // second 0-59 UInt8 hsecond; // 0.01 second 0-99 } </pre>

Name	pString
Element type	18
Element size	1 byte
Description	<p>Pascal string, consisting of a character count (from 0 to 255) in the first byte followed by the 8-bit ASCII characters.</p> <p>Note: Each byte is considered to be an element of the item. The number of elements in the item is equal to the number of characters plus one.</p>

Name	cString
Element type	19
Element size	1 byte
Description	<p>C-style string, consisting of a string of 8-bit ASCII characters followed by a null (zero) byte.</p> <p>Note: Each byte is considered to be an element of the item. The number of elements in the item is equal to the number of characters plus one.</p>

Supported Legacy Data Types

Name	thumb
Element type	12
Element size	10 bytes
Description	<p>The “thumbprint” structure was intended to provide a unique file identifier that could be generated on a local (non-networked) computer and yet would be highly likely to be different from any other thumbprint structure generated on any other computer.</p> <p>It is a packed structure of the following form:</p> <pre> { SInt32 d; SInt32 u; UInt8 c; UInt8 n; } </pre>

Name	bool
Element type	13
Element size	1 byte
Description	One-byte boolean value, with zero meaning false and any other value meaning true.

Name	user
Element type	1024 or greater
Element size	1 byte
Description	A user-defined data structure. The product of the <code>numelements</code> and <code>elementsizes</code> fields may not be equal to the <code>datasize</code> field.

Unsupported Legacy Data Types

Name	rational
Element type	6
Element size	8 bytes
Description	Packed structure to describe an integer fraction: <pre>{ SInt32 numerator; SInt32 denominator }</pre>

Name	BCD
Element type	9
Element size	unknown
Description	Binary-coded decimal value of unknown format.

Name	point
Element type	14
Element size	4 bytes
Description	Packed structure to describe a point, using 16-bit coordinates: <pre>{ SInt16 v; SInt16 h; }</pre>

Name	rect
Element type	15
Element size	8 bytes
Description	Packed structure to describe a rectangle, using 16-bit coordinates: <pre>{ SInt16 top; SInt16 left; SInt16 bottom; SInt16 right; }</pre>

Name	vPoint
Element type	16
Element size	8 bytes
Description	Packed structure to describe a point, using 32-bit coordinates: <pre>{ SInt32 v; SInt32 h; }</pre>

Name	vRect
Element type	17
Element size	16 bytes
Description	<p>Packed structure to describe a rectangle, using 32-bit coordinates:</p> <pre>{ SInt32 top; SInt32 left; SInt32 bottom; SInt32 right; }</pre>

Name	tag
Element type	20
Element size	8 bytes
Description	<p>Packed structure to describe an ABIF tag:</p> <pre>{ SInt32 name; SInt32 number; }</pre>

Name	deltaComp
Element type	128
Element size	
Description	Compressed data.

Name	LZWComp
Element type	256
Element size	
Description	Compressed data.

Name	deltaLZW
Element type	384
Element size	
Description	Compressed data.

Notes on Using the ‘User’ Data Type

Applications have used the `user` data type (element type ≥ 1024) to store opaque OS-specific data structures (such as the Mac OS alias) or compound C or Pascal data structures. This use of `user` type has caused some difficulties in the past when operating systems and/or compilers have changed, due to dependencies on how a particular language and/or compiler represents that data in memory. Your implementations should avoid using the `user` data type in this manner, to eliminate such problems in the future.

When you need to store any user-defined data type, such as a struct or an array of structs, break it into arrays of individual fields, rather than using the `user` data type.

For example, consider the following structure to be stored.

```
struct FooStruct {  
    SInt8 alpha;  
    SInt16 beta;  
    SInt32 gamma;  
} [5]
```

To store this, **DO NOT** define a `user` type and create one data item, as shown below.

Name	FooS
Number	42
Element type	user
Element size	12
Num elements	5

Instead, use three separate data items, one for each field, as shown below:

Name	Alph
Number	42
Element type	byte
Element size	1
Num elements	5

Name	Beta
Number	42
Element type	short
Element size	2
Num elements	5

Name	Gamm
Number	42
Element type	long
Element size	4
Num elements	5

Sample File Schemas

This section contains the tags for ab1 and fsa files generated by the various instrument-software combinations, grouped by instrument and then software version. Optional tags are shown in italic text.

Additional tags that are added to the ab1 files during analysis by SeqScape[®] and/or Sequencing Analysis software are also listed, starting on page 50.

Note: ab1 and fsa files generated by future versions of Data Collection software may contain additional tags or the contents and/or format of the existing tags may be modified.

ABI PRISM® 3100 and 3100-Avant™ Genetic Analyzer Tags

Tags in This Section	Table 1: ab1 File Tags from ABI PRISM® 3100/3100-Avant™ Analyzer Data Collection Software v2.0 on the ABI PRISM® 3100/3100-Avant Genetic Analyzer 23
	Table 2: fsa File Tags from ABI PRISM® 3100/3100-Avant Analyzer Data Collection Software v2.0 on the ABI PRISM 3100/3100-Avant Genetic Analyzer. 26
	Optional tags are shown in italic text.

Table 1 ab1 File Tags from ABI PRISM® 3100/3100-Avant™ Analyzer Data Collection Software v2.0 on the ABI PRISM® 3100/3100-Avant Genetic Analyzer

Name	Number	ABIF Type	Description
APFN	2	pString	Sequencing Analysis parameters file name
APrN	1	cString	Analysis Protocol settings name
APrV	1	cString	Analysis Protocol settings version
APrX	1	char array	Analysis Protocol XML string
APXV	1	cString	Analysis Protocol XML schema version
<i>CMNT</i>	<i>1</i>	<i>pString</i>	<i>Comment about sample (optional)</i>
CpEP	1	byte	Is Capillary Machine?
CTID	1	cString	Container identifier, aka. plate barcode
CTNM	1	cString	Container name (usually identical to CTID)
CTTL	1	pString	Comment title
DATA	1	short array	Channel 1 raw data
DATA	2	short array	Channel 2 raw data
DATA	3	short array	Channel 3 raw data
DATA	4	short array	Channel 4 raw data
DATA	5	short array	Voltage, measured (decaVolts)
DATA	6	short array	Current, measured (milliAmps)
DATA	7	short array	Power, measured (milliWatts)
DATA	8	short array	Temperature, measured (degrees C)
<i>DATA</i>	<i>105</i>	<i>short array</i>	<i>Raw data for dye 5 (optional)</i>
DSam	1	short	Downsampling factor
Dye#	1	short	Number of dyes
DyeN	1	pString	Dye 1 name
DyeN	2	pString	Dye 2 name

Table 1 ab1 File Tags from ABI PRISM® 3100/3100-Avant™ Analyzer Data Collection Software v2.0 on the ABI PRISM® 3100/3100-Avant Genetic Analyzer (continued)

Name	Number	ABIF Type	Description
DyeN	3	pString	Dye 3 name
DyeN	4	pString	Dye 4 name
DyeW	1	short	Dye 1 wavelength
DyeW	2	short	Dye 2 wavelength
DyeW	3	short	Dye 3 wavelength
DyeW	4	short	Dye 4 wavelength
DySN	1	pString	Dye set name
EPVt	1	long	Electrophoresis voltage setting (Volts)
EVNT	1	pString	Start Run event
EVNT	2	pString	Stop Run event
EVNT	3	pString	Start Collection event
EVNT	4	pString	Stop Collection event
FWO_	1	char array	Base order
GTyp	1	pString	Gel type description
InSc	1	long	Injection time (seconds)
InVt	1	long	Injection voltage (Volts)
LANE	1	short	Lane/Capillary
LIMS	1	pString	Sample tracking ID
LNTD	1	short	Length to detector (cm)
LsrP	1	long	Laser Power setting (microWatts)
MCHN	1	pString	Instrument name and serial number
MODF	1	pString	Data collection module file
MODL	1	char[4]	Model number
NAVG	1	short	Pixels averaged per lane
NLNE	1	short	Number of capillaries
OfSc	1	long array	List of scans that are marked off scale in Collection. (optional)
OvrI	1-N	long array	One value for each dye. List of scan number indices for scans with color data values >32767. Values cannot be greater than 32000. (optional)
OvrV	1-N	long array	One value for each dye. List of color data values for the locations listed in the OvrI tag. Number of OvrV tags must be equal to the number of OvrI tags. (optional)
PDMF	1	pString	Mobility file (orig)
PRJT	1	cString	SeqScape® project template name (SeqScape® software sample files only) (optional)
PROJ	4	cString	SeqScape® project name (SeqScape® software sample files only) (optional)

Table 1 ab1 File Tags from ABI PRISM® 3100/3100-Avant™ Analyzer Data Collection Software v2.0 on the ABI PRISM® 3100/3100-Avant Genetic Analyzer (*continued*)

Name	Number	ABIF Type	Description
PXLB	1	long	Pixel bin size
Rate	1	user	Scanning rate
RGCm	1	cString	Results group comment (optional)
RGNm	1	cString	Results group name
RMdN	1	cString	Run Module name (same as MODF)
RMdV	1	cString	Run Module version
RMdX	1	char array	Run Module XML string
RMdXV	1	cString	Run Module XML schema version
RPrN	1	cString	Run Protocol name
RPrV	1	cString	Run Protocol version
RUND	1	date	Run start date
RUND	2	date	Run stop date
RUND	3	date	Data Collection start date
RUND	4	date	Data Collection stop date
RunN	1	cString	Run Name
RUNT	1	time	Run start time
RUNT	2	time	Run stop time
RUNT	3	time	Data Collection start time
RUNT	4	time	Data Collection stop time
Satd	1	long array	Array of longs representing the scan numbers of data points, which are flagged as saturated by data collection (optional)
Scal	1	float	Rescaling divisor for color data
Scan	1	short	Number of scans (legacy - use SCAN)
SCAN	1	long	Number of scans
SMED	1	pString	Polymer lot expiration date
SMLt	1	pString	Polymer lot number
SMPL	1	pString	Sample name
SPEC	1	cString	SeqScape® specimen name (SeqScape® software sample files only) (optional)
SVER	1	pString	Data collection software version
SVER	3	pString	Data collection firmware version
Tmpr	1	long	Run temperature setting (degrees C)
TUBE	1	pString	Well ID
User	1	pString	Name of user who created the plate (optional)

Table 2 fsa File Tags from ABI PRISM® 3100/3100-Avant Analyzer Data Collection Software v2.0 on the ABI PRISM 3100/3100-Avant Genetic Analyzer

Name	Number	ABIF Type	Description
ANME	1	cString	GeneMapper® software analysis method name
CMNT	1-N	pString	<i>Comment(s) about sample (optional)</i>
CpEP	1	byte	Is Capillary Machine?
CTID	1	cString	Container identifier (plate barcode, for example)
CTNM	1	cString	Container name (usually identical to CTID)
CTTL	1	pString	Comment title
DATA	1	short array	Channel 1 raw data
DATA	2	short array	Channel 2 raw data
DATA	3	short array	Channel 3 raw data
DATA	4	short array	Channel 4 raw data
DATA	5	short array	Voltage, measured (decaVolts)
DATA	6	short array	Current, measured (milliAmps)
DATA	7	short array	Power, measured (milliWatts)
DATA	8	short array	Temperature, measured (degrees C)
DATA	105	short array	<i>Raw data for dye 5 (optional)</i>
DSam	1	short	Downsampling factor
Dye#	1	short	Number of dyes
DyeB	1	char	Dye 1 significance S for standard, space for sample
DyeB	2	char	Dye 2 significance S for standard, space for sample
DyeB	3	char	Dye 3 significance S for standard, space for sample
DyeB	4	char	Dye 4 significance S for standard, space for sample
<i>DyeB</i>	5	<i>char</i>	<i>Dye 5 significance S for standard, space for sample (optional)</i>
DyeN	1	pString	Dye 1 name
DyeN	2	pString	Dye 2 name
DyeN	3	pString	Dye 3 name
DyeN	4	pString	Dye 4 name
DyeN	5	pString	<i>Dye 5 name (optional)</i>
DyeW	1	short	Dye 1 wavelength
DyeW	2	short	Dye 2 wavelength
DyeW	3	short	Dye 3 wavelength
DyeW	4	short	Dye 4 wavelength
DyeW	5	short	<i>Dye 5 wavelength (optional)</i>
DySN	1	pString	Dye set name

Table 2 fsa File Tags from ABI PRISM® 3100/3100-Avant Analyzer Data Collection Software v2.0 on the ABI PRISM 3100/3100-Avant Genetic Analyzer (*continued*)

Name	Number	ABIF Type	Description
EPVt	1	long	Electrophoresis voltage setting (Volts)
EVNT	1	pString	Start Run event
EVNT	2	pString	Stop Run event
EVNT	3	pString	Start Collection event
EVNT	4	pString	Stop Collection event
GTyp	1	pString	Gel type description
InSc	1	long	Injection time (seconds)
InVt	1	long	Injection voltage (Volts)
LANE	1	short	Lane/Capillary
LIMS	1	pString	Sample tracking ID
LNTD	1	short	Length to detector (cm)
LsrP	1	long	Laser Power setting (microWatts)
MCHN	1	pString	Instrument name and serial number
MODF	1	pString	Data collection module file
MODL	1	char[4]	Model number
NAVG	1	short	Pixels averaged per lane
NLNE	1	short	Number of capillaries
OfSc	1	long array	List of scans that are marked off scale in Collection. (optional)
Ovrl	1-N	long array	One value for each dye. List of scan number indices for scans with color data values >32767. Values cannot be greater than 32000. (optional)
OvrV	1-N	long array	One value for each dye. List of color data values for the locations listed in the Ovrl tag. Number of OvrV tags must be equal to the number of Ovrl tags. (optional)
PANL	1	cString	GeneMapper® software panel name
PXLB	1	long	Pixel bin size
Rate	1	user	Scan rate.
RGCrn	1	cString	Results group comment (optional)
RGNm	1	cString	Results group name
RMdN	1	cString	Run Module name (same as MODF)
RMdV	1	cString	Run Module version
RMdX	1	char array	Run Module XML string
RMXV	1	cString	Run Module XML schema version
RPrN	1	cString	Run Protocol name
RPrV	1	cString	Run Protocol version
RUND	1	date	Run start date

Table 2 fsa File Tags from ABI PRISM® 3100/3100-Avant Analyzer Data Collection Software v2.0 on the ABI PRISM 3100/3100-Avant Genetic Analyzer (continued)

Name	Number	ABIF Type	Description
RUND	2	date	Run stop date
RUND	3	date	Data Collection start date
RUND	4	date	Data Collection stop date
RunN	1	cString	Run Name
RUNT	1	time	Run start time
RUNT	2	time	Run stop time
RUNT	3	time	Data Collection start time
RUNT	4	time	Data Collection stop time
<i>Satd</i>	<i>1</i>	<i>long array</i>	<i>Array of longs representing the scan numbers of data points, which are flagged as saturated by data collection (optional)</i>
Scal	1	float	Rescaling divisor for color data
Scan	1	short	Number of scans (legacy - use SCAN)
SCAN	1	long	Number of scans
SMED	1	pString	Polymer lot expiration date
SMLt	1	pString	Polymer lot number
<i>SnpS</i>	<i>1</i>	<i>pString</i>	<i>SNP set name (optional)</i>
SpNm	1	pString	Sample name for GeneScan® software sample files
StdF	1	pString	Size Standard file name
STYP	1	cString	GeneMapper® software Sample Type
SVER	1	pString	Data collection software version
SVER	3	pString	Data collection firmware version
SVER	4	pString	Sample File Format Version, containing the version of the sample file format used to write the file
Tmpr	1	long	Run temperature setting (degrees C)
TUBE	1	pString	Well ID
<i>User</i>	<i>1</i>	<i>pString</i>	<i>Name of user who created the plate (optional)</i>

Applied Biosystems 3130/3130xl Genetic Analyzer Tags

Tags in This Section	Table 3: ab1 File Tags from Applied Biosystems 3130/3130xl Data Collection Software v3.0 on the Applied Biosystems 3130/3130xl Genetic Analyzer.	29
	Table 4: fsa File Tags from Applied Biosystems 3130/3130xl Data Collection Software v3.0 on the Applied Biosystems 3130/3130xl Genetic Analyzer.	33

Optional tags are shown in italic text.

Table 3 ab1 File Tags from Applied Biosystems 3130/3130xl Data Collection Software v3.0 on the Applied Biosystems 3130/3130xl Genetic Analyzer

Name	Number	ABIF Type	Description
APFN	2	pString	Sequencing Analysis parameters file name
APrN	1	cString	Analysis Protocol settings name
APrV	1	cString	Analysis Protocol settings version
APrX	1	char array	Analysis Protocol XML string
APXV	1	cString	Analysis Protocol XML schema version
<i>CMNT</i>	<i>1</i>	<i>pString</i>	<i>Comment about sample (optional)</i>
CpEP	1	byte	Is Capillary Machine?
CTID	1	cString	Container identifier (plate barcode, for example)
CTNM	1	cString	Container name (usually identical to CTID)
CTOw	1	cString	Container owner
CTTL	1	pString	Comment title
DATA	1	short array	Channel 1 raw data
DATA	2	short array	Channel 2 raw data
DATA	3	short array	Channel 3 raw data
DATA	4	short array	Channel 4 raw data
DATA	5	short array	Voltage, measured (decaVolts)
DATA	6	short array	Current, measured (milliAmps)
DATA	7	short array	Power, measured (milliWatts)
DATA	8	short array	Temperature, measured (degrees C)
<i>DATA</i>	<i>105</i>	<i>short array</i>	<i>Raw data for dye 5 (optional)</i>
DSam	1	short	Downsampling factor
Dye#	1	short	Number of dyes
DyeN	1	pString	Dye 1 name

Table 3 ab1 File Tags from Applied Biosystems 3130/3130x/ Data Collection Software v3.0 on the Applied Biosystems 3130/3130x/ Genetic Analyzer (continued)

Name	Number	ABIF Type	Description
DyeN	2	pString	Dye 2 name
DyeN	3	pString	Dye 3 name
DyeN	4	pString	Dye 4 name
DyeW	1	short	Dye 1 wavelength
DyeW	2	short	Dye 2 wavelength
DyeW	3	short	Dye 3 wavelength
DyeW	4	short	Dye 4 wavelength
DySN	1	pString	Dye set name
EPVt	1	long	Electrophoresis voltage setting (Volts)
EVNT	1	pString	Start Run event
EVNT	2	pString	Stop Run event
EVNT	3	pString	Start Collection event
EVNT	4	pString	Stop Collection event
FWO_	1	char array	Base order
GTyp	1	pString	Gel type description
HCFG	1	cString	Instrument Class
HCFG	2	cString	Instrument Family
HCFG	3	cString	Official Instrument Name
HCFG	4	cString	Instrument Parameters
InSc	1	long	Injection time (seconds)
InVt	1	long	Injection voltage (Volts)
LANE	1	short	Lane/Capillary
LIMS	1	pString	Sample tracking ID
LNTD	1	short	Length to detector (cm)
LsrP	1	long	Laser Power setting (microWatts)
MCHN	1	pString	Instrument name and serial number
MODF	1	pString	Data collection module file
MODL	1	char[4]	Model number
NAVG	1	short	Pixels averaged per lane
NLNE	1	short	Number of capillaries
<i>OfSc</i>	<i>1</i>	<i>long array</i>	<i>List of scans that are marked off scale in Collection. (optional)</i>
<i>OvrI</i>	<i>1-N</i>	<i>long array</i>	<i>One value for each dye. List of scan number indices for scans with color data values >32767. Values cannot be greater than 32000. (optional)</i>
<i>OvrV</i>	<i>1-N</i>	<i>long array</i>	<i>One value for each dye. List of color data values for the locations listed in the OvrI tag. Number of OvrV tags must be equal to the number of OvrI tags. (optional)</i>

Table 3 ab1 File Tags from Applied Biosystems 3130/3130xl Data Collection Software v3.0 on the Applied Biosystems 3130/3130xl Genetic Analyzer (*continued*)

Name	Number	ABIF Type	Description
PDMF	1	pString	Mobility file (orig)
PRJT	1	cString	SeqScape® project template name (SeqScape® software sample files only) (optional)
PROJ	4	cString	SeqScape® project name (SeqScape® software sample files only) (optional)
PXLB	1	long	Pixel bin size
Rate	1	user	Scanning rate
RGCm	1	cString	Results group comment (optional)
RGNm	1	cString	Results group name
RGOw	1	cString	The name entered as the owner of a results group, in the Results Group editor (optional)
RMdN	1	cString	Run Module name (same as MODF)
RMdV	a1	cString	Run Module version
RMdX	1	char array	Run Module XML string
RMXV	1	cString	Run Module XML schema version
RPrN	1	cString	Run Protocol name
RPrV	1	cString	Run Protocol version
RUND	1	date	Run start date
RUND	2	date	Run stop date
RUND	3	date	Data Collection start date
RUND	4	date	Data Collection stop date
RunN	1	cString	Run Name
RUNT	1	time	Run start time
RUNT	2	time	Run stop time
RUNT	3	time	Data Collection start time
RUNT	4	time	Data Collection stop time
Satd	1	long array	Array of longs representing the scan numbers of data points, which are flagged as saturated by data collection (optional)
Scal	1	float	Rescaling divisor for color data
Scan	1	short	Number of scans (legacy - use SCAN)
SCAN	1	long	Number of scans
SMED	1	pString	Polymer lot expiration date
SMLt	1	pString	Polymer lot number
SMPL	1	pString	Sample name
SPEC	1	cString	SeqScape® specimen name (SeqScape® software sample files only) (optional)

Table 3 ab1 File Tags from Applied Biosystems 3130/3130x/ Data Collection Software v3.0 on the Applied Biosystems 3130/3130x/ Genetic Analyzer (*continued*)

Name	Number	ABIF Type	Description
SVER	1	pString	Data collection software version
SVER	3	pString	Data collection firmware version
Tmpr	1	long	Run temperature setting (degrees C)
TUBE	1	pString	Well ID
<i>User</i>	<i>1</i>	<i>pString</i>	<i>Name of user who created the plate (optional)</i>

Table 4 fsa File Tags from Applied Biosystems 3130/3130xl Data Collection Software v3.0 on the Applied Biosystems 3130/3130xl Genetic Analyzer

Name	Number	ABIF Type	Description
CMNT	1-N	pString	Comment(s) about sample (optional)
CpEP	1	byte	Is Capillary Machine?
CTID	1	cString	Container identifier (plate barcode, for example)
CTNM	1	cString	Container name (usually identical to CTID)
CTOw	1	cString	Container owner
CTTL	1	pString	Comment title
DATA	1	short array	Channel 1 raw data
DATA	2	short array	Channel 2 raw data
DATA	3	short array	Channel 3 raw data
DATA	4	short array	Channel 4 raw data
DATA	5	short array	Voltage, measured (decaVolts)
DATA	6	short array	Current, measured (milliAmps)
DATA	7	short array	Power, measured (milliWatts)
DATA	8	short array	Temperature, measured (degrees C)
DATA	105	short array	Raw data for dye 5 (optional)
DCHT	1	short	Detector cell heater temperature (degrees C)
DSam	1	short	Downsampling factor
Dye#	1	short	Number of dyes
DyeB	1	char	Dye 1 significance S for standard, space for sample
DyeB	2	char	Dye 2 significance S for standard, space for sample
DyeB	3	char	Dye 3 significance S for standard, space for sample
DyeB	4	char	Dye 4 significance S for standard, space for sample
DyeB	5	char	Dye 5 significance S for standard, space for sample (optional)
DyeN	1	pString	Dye 1 name
DyeN	2	pString	Dye 2 name
DyeN	3	pString	Dye 3 name
DyeN	4	pString	Dye 4 name
DyeN	5	pString	Dye 5 name (optional)
DyeW	1	short	Dye 1 wavelength
DyeW	2	short	Dye 2 wavelength
DyeW	3	short	Dye 3 wavelength
DyeW	4	short	Dye 4 wavelength
DyeW	5	short	Dye 5 wavelength (optional)

Table 4 fsa File Tags from Applied Biosystems 3130/3130x/ Data Collection Software v3.0 on the Applied Biosystems 3130/3130x/ Genetic Analyzer (continued)

Name	Number	ABIF Type	Description
DySN	1	pString	Dye set name
EPVt	1	long	Electrophoresis voltage setting (Volts)
EVNT	1	pString	Start Run event
EVNT	2	pString	Stop Run event
EVNT	3	pString	Start Collection event
EVNT	4	pString	Stop Collection event
GTyp	1	pString	Gel type description
HCFG	1	cString	Instrument Class
HCFG	2	cString	Instrument Family
HCFG	3	cString	Official Instrument Name
HCFG	4	cString	Instrument Parameters
InSc	1	long	Injection time (seconds)
InVt	1	long	Injection voltage (Volts)
LANE	1	short	Lane/Capillary
LIMS	1	pString	Sample tracking ID
LNTD	1	short	Length to detector (cm)
LsrP	1	long	Laser Power setting (microWatts)
MCHN	1	pString	Instrument name and serial number
MODF	1	pString	Data collection module file
MODL	1	char[4]	Model number
NAVG	1	short	Pixels averaged per lane
NLNE	1	short	Number of capillaries
OffS	1-N	user	Range of offscale peaks (optional)
OfSc	1	long array	List of scans that are marked off scale in Collection. (optional)
OvrI	1-N	long array	One value for each dye. List of scan number indices for scans with color data values >32767. Values cannot be greater than 32000. (optional)
OvrV	1-N	long array	One value for each dye. List of color data values for the locations listed in the OvrI tag. Number of OvrV tags must be equal to the number of OvrI tags. (optional)
PANL	1	cString	GeneMapper® software panel name
PSZE	1	long	Plate size. The number of sample positions in the container; allowed values 96 and 384
PTYP	1	cString	Plate type. Allowed values are 96-Well, 384-Well.
PXLB	1	long	Pixel bin size
Rate	1	user	Scan rate

Table 4 fsa File Tags from Applied Biosystems 3130/3130xl Data Collection Software v3.0 on the Applied Biosystems 3130/3130xl Genetic Analyzer (continued)

Name	Number	ABIF Type	Description
<i>RGCm</i>	1	<i>cString</i>	<i>Results group comment (optional)</i>
RGNm	1	cString	Results group name
<i>RGOw</i>	1	<i>cString</i>	<i>The name entered as the owner of a results group, in the Results Group editor (optional)</i>
RMdN	1	cString	Run Module name (same as MODF)
RMdV	1	cString	Run Module version
RMdX	1	char array	Run Module XML string
RMXV	1	cString	Run Module XML schema version
RPrN	1	cString	Run Protocol name
RPrV	1	cString	Run Protocol version
RUND	1	date	Run start date
RUND	2	date	Run stop date
RUND	3	date	Data Collection start date
RUND	4	date	Data Collection stop date
RunN	1	cString	Run Name
RUNT	1	time	Run start time
RUNT	2	time	Run stop time
RUNT	3	time	Data Collection start time
RUNT	4	time	Data Collection stop time
<i>Satd</i>	1	<i>long array</i>	<i>Array of longs representing the scan numbers of data points, which are flagged as saturated by data collection (optional)</i>
Scal	1	float	Rescaling divisor for color data
Scan	1	short	Number of scans (legacy - use SCAN)
SCAN	1	long	Number of scans
SMED	1	pString	Polymer lot expiration date
SMLt	1	pString	Polymer lot number
<i>SnpS</i>	1	<i>pString</i>	<i>SNP set name (optional)</i>
SpNm	1	pString	Sample name for GeneScan® software sample files
SVER	1	pString	Data collection software version
SVER	3	pString	Data collection firmware version
SVER	4	pString	Sample File Format Version, containing the version of the sample file format used to write the file
Tmpr	1	long	Run temperature setting (degrees C)
TUBE	1	pString	Well ID
<i>User</i>	1	<i>pString</i>	<i>Name of user who created the plate (optional)</i>

Applied Biosystems 3730/3730x/ DNA Analyzer Tags

Tags in This Section	Table 5: ab1 File Tags from Applied Biosystems 3730/3730xl Data Collection Software v2.0 on the Applied Biosystems 3730/3730xl Genetic Analyzer36
	Table 6: fsa File Tags from Applied Biosystems 3730/3730xl Data Collection Software v2.0 on the Applied Biosystems 3730/3730xl DNA Analyzer40
	Table 7: ab1 File Tags from Applied Biosystems 3730/3730xl Data Collection Software v3.0 on the Applied Biosystems 3730/3730xl DNA Analyzer DNA Analyzer.....43
	Table 8: fsa File Tags from Applied Biosystems 3730/3730xl Data Collection Software v3.0 on the Applied Biosystems 3730/3730xl DNA Analyzer46

Optional tags are shown in italic text.

Table 5 ab1 File Tags from Applied Biosystems 3730/3730x/ Data Collection Software v2.0 on the Applied Biosystems 3730/3730x/ Genetic Analyzer

Name	Number	ABIF Type	Description
APFN	2	pString	Sequencing Analysis parameters file name
APrN	1	cString	Analysis Protocol settings name
APrV	1	cString	Analysis Protocol settings version
APrX	1	char array	Analysis Protocol XML string
APXV	1	cString	Analysis Protocol XML schema version
BuT	1	short array	Buffer tray heater temperature (degrees C)
<i>CMNT</i>	<i>1</i>	<i>pString</i>	<i>Comment about sample (optional)</i>
CpEP	1	byte	Is Capillary Machine?
CTID	1	cString	Container identifier (plate barcode, for example)
CTNM	1	cString	Container name (usually identical to CTID)
C TTL	1	pString	Comment title
DATA	1	short array	Channel 1 raw data
DATA	2	short array	Channel 2 raw data
DATA	3	short array	Channel 3 raw data
DATA	4	short array	Channel 4 raw data
DATA	5	short array	Voltage, measured (decaVolts)
DATA	6	short array	Current, measured (milliAmps)

Table 5 ab1 File Tags from Applied Biosystems 3730/3730xl Data Collection Software v2.0 on the Applied Biosystems 3730/3730xl Genetic Analyzer (continued)

Name	Number	ABIF Type	Description
DATA	7	short array	Power, measured (milliWatts)
DATA	8	short array	Temperature, measured (degrees C)
DATA	105	short array	Raw data for dye 5 (optional)
DSam	1	short	Downsampling factor
Dye#	1	short	Number of dyes
DyeN	1	pString	Dye 1 name
DyeN	2	pString	Dye 2 name
DyeN	3	pString	Dye 3 name
DyeN	4	pString	Dye 4 name
DyeN	5	pString	Dye 5 name (optional)
DyeW	1	short	Dye 1 wavelength
DyeW	2	short	Dye 2 wavelength
DyeW	3	short	Dye 3 wavelength
DyeW	4	short	Dye 4 wavelength
DyeW	5	short	Dye 5 wavelength (optional)
DySN	1	pString	Dye set name
EPVt	1	long	Electrophoresis voltage setting (Volts)
EVNT	1	pString	Start Run event
EVNT	2	pString	Stop Run event
EVNT	3	pString	Start Collection event
EVNT	4	pString	Stop Collection event
FWO_	1	char array	Base order
GTyp	1	pString	Gel type description
InSc	1	long	Injection time (seconds)
InVt	1	long	Injection voltage (Volts)
LANE	1	short	Lane/Capillary
LIMS	1	pString	Sample tracking ID
LNTD	1	short	Length to detector (cm)
LsrP	1	long	Laser Power setting (microWatts)
MCHN	1	pString	Instrument name and serial number
MODF	1	pString	Data collection module file
MODL	1	char[4]	Model number
NAVG	1	short	Pixels averaged per lane
NLNE	1	short	Number of capillaries

Table 5 ab1 File Tags from Applied Biosystems 3730/3730x/ Data Collection Software v2.0 on the Applied Biosystems 3730/3730x/ Genetic Analyzer (continued)

Name	Number	ABIF Type	Description
<i>OfSc</i>	1	long array	List of scans that are marked off scale in Collection. (optional)
<i>Ovrl</i>	1-N	long array	One value for each dye. List of scan number indices for scans with color data values >32767. Values cannot be greater than 32000. (optional)
<i>OvrV</i>	1-N	long array	One value for each dye. List of color data values for the locations listed in the <i>Ovrl</i> tag. Number of <i>OvrV</i> tags must be equal to the number of <i>Ovrl</i> tags. (optional)
PDMF	1	pString	Mobility file (orig)
<i>PRJT</i>	1	cString	SeqScape® project template name (SeqScape® software sample files only) (optional)
<i>PROJ</i>	4	cString	SeqScape® project name (SeqScape® software sample files only) (optional)
PXLB	1	long	Pixel bin size
Rate	1	user	Scan rate.
RGCM	1	cString	Results group comment (optional)
RGNM	1	cString	Results group name
RMdN	1	cString	Run Module name (same as MODF)
RMdV	1	cString	Run Module version
RMdX	1	char array	Run Module XML string
RMXV	1	cString	Run Module XML schema version
RPrN	1	cString	Run Protocol name
RPrV	1	cString	Run Protocol version
RUND	1	date	Run start date
RUND	2	date	Run stop date
RUND	3	date	Data Collection start date
RUND	4	date	Data Collection stop date
RunN	1	cString	Run Name
RUNT	1	time	Run start time
RUNT	2	time	Run stop time
RUNT	3	time	Data Collection start time
RUNT	4	time	Data Collection stop time
<i>Satd</i>	1	long array	Array of longs representing the scan numbers of data points, which are flagged as saturated by data collection (optional)
Scal	1	float	Rescaling divisor for color data
Scan	1	short	Number of scans (legacy - use SCAN)
SCAN	1	long	Number of scans
SMED	1	pString	Polymer lot expiration date

Table 5 ab1 File Tags from Applied Biosystems 3730/3730xl Data Collection Software v2.0 on the Applied Biosystems 3730/3730xl Genetic Analyzer (continued)

Name	Number	ABIF Type	Description
SMLt	1	pString	Polymer lot number
SMPL	1	pString	Sample name
SPEC	1	cString	SeqScape [®] specimen name (SeqScape [®] software sample files only) (optional)
SVER	1	pString	Data collection software version
SVER	3	pString	Data collection firmware version
Tmpr	1	long	Run temperature setting (degrees C)
TUBE	1	pString	Well ID
User	1	pString	Name of user who created the plate (optional)

Table 6 fsa File Tags from Applied Biosystems 3730/3730x/ Data Collection Software v2.0 on the Applied Biosystems 3730/3730x/ DNA Analyzer

Name	Number	ABIF Type	Description
ANME	1	cString	GeneMapper® software analysis method name
BuFT	1	short array	Buffer tray heater temperature (degrees C)
CMNT	1-N	pString	Comment(s) about sample (optional)
CpEP	1	byte	Is Capillary Machine?
CTID	1	cString	Container identifier (plate barcode, for example)
CTNM	1	cString	Container name (usually identical to CTID)
CTTL	1	pString	Comment title
DATA	1	short array	Channel 1 raw data
DATA	2	short array	Channel 2 raw data
DATA	3	short array	Channel 3 raw data
DATA	4	short array	Channel 4 raw data
DATA	5	short array	Voltage, measured (decaVolts)
DATA	6	short array	Current, measured (milliAmps)
DATA	7	short array	Power, measured (milliWatts)
DATA	8	short array	Temperature, measured (degrees C)
DATA	105	short array	Raw data for dye 5 (optional)
DSam	1	short	Downsampling factor
Dye#	1	short	Number of dyes
DyeB	1	char	Dye 1 significance S for standard, space for sample
DyeB	2	char	Dye 2 significance S for standard, space for sample
DyeB	3	char	Dye 3 significance S for standard, space for sample
DyeB	4	char	Dye 4 significance S for standard, space for sample
DyeB	5	char	Dye 5 significance S for standard, space for sample (optional)
DyeN	1	pString	Dye 1 name
DyeN	2	pString	Dye 2 name
DyeN	3	pString	Dye 3 name
DyeN	4	pString	Dye 4 name
DyeN	5	pString	Dye 5 name
DyeW	1	short	Dye 1 wavelength
DyeW	2	short	Dye 2 wavelength
DyeW	3	short	Dye 3 wavelength
DyeW	4	short	Dye 4 wavelength
DyeW	5	short	Dye 5 wavelength (optional)

Table 6 fsa File Tags from Applied Biosystems 3730/3730xl Data Collection Software v2.0 on the Applied Biosystems 3730/3730xl DNA Analyzer (*continued*)

Name	Number	ABIF Type	Description
DySN	1	pString	Dye set name
EPVt	1	long	Electrophoresis voltage setting (Volts)
EVNT	1	pString	Start Run event
EVNT	2	pString	Stop Run event
EVNT	3	pString	Start Collection event
EVNT	4	pString	Stop Collection event
GTyp	1	pString	Gel type description
InSc	1	long	Injection time (seconds)
InVt	1	long	Injection voltage (Volts)
LANE	1	short	Lane/Capillary
LIMS	1	pString	Sample tracking ID
LNTD	1	short	Length to detector (cm)
LsrP	1	long	Laser Power setting (microWatts)
MCHN	1	pString	Instrument name and serial number
MODF	1	pString	Data collection module file
MODL	1	char[4]	Model number
NAVG	1	short	Pixels averaged per lane
NLNE	1	short	Number of capillaries
OffS	1-N	user	Range of offscale peaks (<i>optional</i>)
OfSc	1	long array	List of scans that are marked off scale in Collection. (<i>optional</i>)
OvrI	1-N	long array	One value for each dye. List of scan number indices for scans with color data values >32767. Values cannot be greater than 32000. (<i>optional</i>)
OvrV	1-N	long array	One value for each dye. List of color data values for the locations listed in the OvrI tag. Number of OvrV tags must be equal to the number of OvrI tags. (<i>optional</i>)
PANL	1	cString	GeneMapper® software panel name
PXLB	1	long	Pixel bin size
Rate	1	user	Scan rate.
RGCm	1	cString	Results group comment (<i>optional</i>)
RGNm	1	cString	Results group name
RMdN	1	cString	Run Module name (same as MODF)
RMdV	1	cString	Run Module version
RMdX	1	char array	Run Module XML string
RMXV	1	cString	Run Module XML schema version
RPrN	1	cString	Run Protocol name

Table 6 fsa File Tags from Applied Biosystems 3730/3730xl Data Collection Software v2.0 on the Applied Biosystems 3730/3730xl DNA Analyzer (continued)

Name	Number	ABIF Type	Description
RPv	1	cString	Run Protocol version
RUND	1	date	Run start date
RUND	2	date	Run stop date
RUND	3	date	Data Collection start date
RUND	4	date	Data Collection stop date
RunN	1	cString	Run Name
RUNT	1	time	Run start time
RUNT	2	time	Run stop time
RUNT	3	time	Data Collection start time
RUNT	4	time	Data Collection stop time
<i>Satd</i>	<i>1</i>	<i>long array</i>	<i>Array of longs representing the scan numbers of data points, which are flagged as saturated by data collection (optional)</i>
Scal	1	float	Rescaling divisor for color data
Scan	1	short	Number of scans (legacy - use SCAN)
SCAN	1	long	Number of scans
SMED	1	pString	Polymer lot expiration date
SMLt	1	pString	Polymer lot number
<i>SnpS</i>	<i>1</i>	<i>pString</i>	<i>SNP set name (optional)</i>
SpNm	1	pString	Sample name for GeneScan® software sample files
StdF	1	pString	Size Standard file name
STYP	1	cString	GeneMapper® software Sample Type
SVER	1	pString	Data collection software version
SVER	3	pString	Data collection firmware version
SVER	4	pString	Sample File Format Version, containing the version of the sample file format used to write the file
Tmpr	1	long	Run temperature setting (degrees C)
TUBE	1	pString	Well ID
<i>User</i>	<i>1</i>	<i>pString</i>	<i>Name of user who created the plate (optional)</i>

Table 7 ab1 File Tags from Applied Biosystems 3730/3730xl Data Collection Software v3.0 on the Applied Biosystems 3730/3730xl DNA Analyzer DNA Analyzer

Name	Number	ABIF Type	Description
APFN	2	pString	Sequencing Analysis parameters file name
APrN	1	cString	Analysis Protocol settings name
APrV	1	cString	Analysis Protocol settings version
APrX	1	char array	Analysis Protocol XML string
APXV	1	cString	Analysis Protocol XML schema version
BufT	1	short array	Buffer tray heater temperature (degrees C)
CMNT	1	pString	<i>Comment about sample (optional)</i>
CpEP	1	byte	Is Capillary Machine?
CTID	1	cString	Container identifier (plate barcode, for example)
CTNM	1	cString	Container name (usually identical to CTID)
CTOw	1	cString	Container owner
CTTL	1	pString	Comment title
DATA	1	short array	Channel 1 raw data
DATA	2	short array	Channel 2 raw data
DATA	3	short array	Channel 3 raw data
DATA	4	short array	Channel 4 raw data
DATA	5	short array	Voltage, measured (decaVolts)
DATA	6	short array	Current, measured (milliAmps)
DATA	7	short array	Power, measured (milliWatts)
DATA	8	short array	Temperature, measured (degrees C)
DATA	105	short array	<i>Raw data for dye 5 (optional)</i>
DCHT	1	short	Detector cell heater temperature (degrees C)
DSam	1	short	Downsampling factor
Dye#	1	short	Number of dyes
DyeN	1	pString	Dye 1 name
DyeN	2	pString	Dye 2 name
DyeN	3	pString	Dye 3 name
DyeN	4	pString	Dye 4 name
DyeW	1	short	Dye 1 wavelength
DyeW	2	short	Dye 2 wavelength
DyeW	3	short	Dye 3 wavelength
DyeW	4	short	Dye 4 wavelength
DySN	1	pString	Dye set name

Table 7 ab1 File Tags from Applied Biosystems 3730/3730x/ Data Collection Software v3.0 on the Applied Biosystems 3730/3730x/ DNA Analyzer DNA Analyzer (continued)

Name	Number	ABIF Type	Description
EPVt	1	long	Electrophoresis voltage setting (Volts)
EVNT	1	pString	Start Run event
EVNT	2	pString	Stop Run event
EVNT	3	pString	Start Collection event
EVNT	4	pString	Stop Collection event
FWO_	1	char array	Base order
GTyp	1	pString	Gel type description
HCFG	1	cString	Instrument Class
HCFG	2	cString	Instrument Family
HCFG	3	cString	Official Instrument Name
HCFG	4	cString	Instrument Parameters
InSc	1	long	Injection time (seconds)
InVt	1	long	Injection voltage (Volts)
LANE	1	short	Lane/Capillary
LIMS	1	pString	Sample tracking ID
LNTD	1	short	Length to detector (cm)
LsrP	1	long	Laser Power setting (microWatts)
MCHN	1	pString	Instrument name and serial number
MODF	1	pString	Data collection module file
MODL	1	char[4]	Model number
NAVG	1	short	Pixels averaged per lane
NLNE	1	short	Number of capillaries
OfSc	1	long array	List of scans that are marked off scale in Collection. (optional)
OvrI	1-N	long array	One value for each dye. List of scan number indices for scans with color data values >32767. Values cannot be greater than 32000. (optional)
OvrV	1-N	long array	One value for each dye. List of color data values for the locations listed in the OvrI tag. Number of OvrV tags must be equal to the number of OvrI tags. (optional)
PDMF	1	pString	Mobility file (orig)
PRJT	1	cString	SeqScape® project template name (SeqScape® software sample files only) (optional)
PROJ	4	cString	SeqScape® project template name (SeqScape® software sample files only) (optional)
PSZE	1	long	Plate size. The number of sample positions in the container; allowed values 96 and 384
PTYP	1	cString	Plate type. Allowed values are 96-Well, 384-Well.

Table 7 ab1 File Tags from Applied Biosystems 3730/3730xl Data Collection Software v3.0 on the Applied Biosystems 3730/3730xl DNA Analyzer DNA Analyzer (continued)

Name	Number	ABIF Type	Description
PXLB	1	long	Pixel bin size
Rate	1	user	Scan rate.
<i>RGCm</i>	<i>1</i>	<i>cString</i>	<i>Results group comment (optional)</i>
RGNm	1	cString	Results group name
<i>RGOw</i>	<i>1</i>	<i>cString</i>	<i>The name entered as the owner of a results group, in the Results Group editor (optional)</i>
RMdN	1	cString	Run Module name (same as MODF)
RMdV	1	cString	Run Module version
RMdX	1	char array	Run Module XML string
RMXV	1	cString	Run Module XML schema version
RPrN	1	cString	Run Protocol name
RPrV	1	cString	Run Protocol version
RUND	1	date	Run start date
RUND	2	date	Run stop date
RUND	3	date	Data Collection start date
RUND	4	date	Data Collection stop date
RunN	1	cString	Run Name
RUNT	1	time	Run start time
RUNT	2	time	Run stop time
RUNT	3	time	Data Collection start time
RUNT	4	time	Data Collection stop time
<i>Satd</i>	<i>1</i>	<i>long array</i>	<i>Array of longs representing the scan numbers of data points, which are flagged as saturated by data collection (optional)</i>
Scal	1	float	Rescaling divisor for color data
Scan	1	short	Number of scans (legacy - use SCAN)
SCAN	1	long	Number of scans
SMED	1	pString	Polymer lot expiration date
SMLt	1	pString	Polymer lot number
SMPL	1	pString	Sample name
<i>SPEC</i>	<i>1</i>	<i>cString</i>	<i>SeqScape® specimen name (SeqScape® software sample files only) (optional)</i>
SVER	1	pString	Data collection software version
SVER	3	pString	Data collection firmware version
Tmpr	1	long	Run temperature setting (degrees C)
TUBE	1	pString	Well ID
<i>User</i>	<i>1</i>	<i>pString</i>	<i>Name of user who created the plate (optional)</i>

Table 8 fsa File Tags from Applied Biosystems 3730/3730x/ Data Collection Software v3.0 on the Applied Biosystems 3730/3730x/ DNA Analyzer

Name	Number	ABIF Type	Description
ANME	1	cString	GeneMapper® analysis method name
BuT	1	short array	Buffer tray heater temperature (degrees C)
CMNT	1-N	pString	<i>Comment about sample (optional)</i>
CpEP	1	byte	Is Capillary Machine?
CTID	1	cString	Container identifier, aka. plate barcode
CTNM	1	cString	Container name (usually identical to CTID)
CTOW	1	cString	Container owner
CTTL	1	pString	Comment title
DATA	1	short array	Channel 1 raw data
DATA	2	short array	Channel 2 raw data
DATA	3	short array	Channel 3 raw data
DATA	4	short array	Channel 4 raw data
DATA	5	short array	Voltage, measured (decaVolts)
DATA	6	short array	Current, measured (milliAmps)
DATA	7	short array	Power, measured (milliWatts)
DATA	8	short array	Temperature, measured (degrees C)
DATA	105	short array	<i>Raw data for dye 5 (optional)</i>
DCHT	1	short	Detector cell heater temperature (degrees C)
DSam	1	short	Downsampling factor
Dye#	1	short	Number of dyes
DyeB	1	char	Dye 1 significance S for standard, space for sample
DyeB	2	char	Dye 2 significance S for standard, space for sample
DyeB	3	char	Dye 3 significance S for standard, space for sample
DyeB	4	char	Dye 4 significance S for standard, space for sample
DyeB	5	char	<i>Dye 5 significance S for standard, space for sample (optional)</i>
DyeN	1	pString	Dye 1 name
DyeN	2	pString	Dye 2 name
DyeN	3	pString	Dye 3 name
DyeN	4	pString	Dye 4 name
DyeN	5	pString	<i>Dye 5 name (optional)</i>
DyeW	1	short	Dye 1 wavelength
DyeW	2	short	Dye 2 wavelength
DyeW	3	short	Dye 3 wavelength

Table 8 fsa File Tags from Applied Biosystems 3730/3730xl Data Collection Software v3.0 on the Applied Biosystems 3730/3730xl DNA Analyzer (*continued*)

Name	Number	ABIF Type	Description
DyeW	4	short	Dye 4 wavelength
<i>DyeW</i>	5	<i>short</i>	<i>Dye 5 wavelength (optional)</i>
DySN	1	pString	Dye set name
EPVt	1	long	Electrophoresis voltage setting (Volts)
EVNT	1	pString	Start Run event
EVNT	2	pString	Stop Run event
EVNT	3	pString	Start Collection event
EVNT	4	pString	Stop Collection event
GTyp	1	pString	Gel type description
HCFG	1	cString	Instrument Class
HCFG	2	cString	Instrument Family
HCFG	3	cString	Official Instrument Name
HCFG	4	cString	Instrument Parameters
InSc	1	long	Injection time (seconds)
InVt	1	long	Injection voltage (Volts)
LANE	1	short	Lane/Capillary
LIMS	1	pString	Sample tracking ID
LNTD	1	short	Length to detector (cm)
LsrP	1	long	Laser Power setting (microWatts)
MCHN	1	pString	Instrument name and serial number
MODF	1	pString	Data collection module file
MODL	1	char[4]	Model number
NAVG	1	short	Pixels averaged per lane
NLNE	1	short	Number of capillaries
<i>OffS</i>	<i>1-N</i>	<i>user</i>	<i>Range of offscale peaks (optional)</i>
<i>OfSc</i>	<i>1</i>	<i>long array</i>	<i>List of scans that are marked off scale in Collection. (optional)</i>
<i>OvrI</i>	<i>1-N</i>	<i>long array</i>	<i>One value for each dye. List of scan number indices for scans with color data values >32767. Values cannot be greater than 32000. (optional)</i>
<i>OvrV</i>	<i>1-N</i>	<i>long array</i>	<i>One value for each dye. List of color data values for the locations listed in the OvrI tag. Number of OvrV tags must be equal to the number of OvrI tags. (optional)</i>
PANL	1	cString	GeneMapper® panel name
PSZE	1	long	Plate size. The number of sample positions in the container; allowed values 96 and 384
PTYP	1	cString	Plate type. Allowed values are 96-Well, 384-Well.

Table 8 fsa File Tags from Applied Biosystems 3730/3730xl Data Collection Software v3.0 on the Applied Biosystems 3730/3730xl DNA Analyzer (continued)

Name	Number	ABIF Type	Description
PXLB	1	long	Pixel bin size
Rate	1	user	Scan rate.
<i>RGCM</i>	<i>1</i>	<i>cString</i>	<i>Results group comment (optional)</i>
RGNm	1	cString	Results group name
<i>RGOW</i>	<i>1</i>	<i>cString</i>	<i>The name entered as the owner of a results group, in the Results Group editor (optional)</i>
RMdN	1	cString	Run Module name (same as MODF)
RMdV	1	cString	Run Module version
RMdX	1	char array	Run Module XML string
RMXV	1	cString	Run Module XML schema version
RPrN	1	cString	Run Protocol name
RPrV	1	cString	Run Protocol version
RUND	1	date	Run start date
RUND	2	date	Run stop date
RUND	3	date	Data Collection start date
RUND	4	date	Data Collection stop date
RunN	1	cString	Run Name
RUNT	1	time	Run start time
RUNT	2	time	Run stop time
RUNT	3	time	Data Collection start time
RUNT	4	time	Data Collection stop time
<i>Satd</i>	<i>1</i>	<i>long array</i>	<i>Array of longs representing the scan numbers of data points which are flagged as saturated by data collection (optional)</i>
Scal	1	float	Rescaling divisor for color data
Scan	1	short	Number of scans (legacy - use SCAN)
SCAN	1	long	Number of scans
SMED	1	pString	Polymer lot expiration date
SMLt	1	pString	Polymer lot number
<i>SnpS</i>	<i>1</i>	<i>pString</i>	<i>SNP set name (optional)</i>
SpNm	1	pString	Sample name for GeneScan [®] sample files
StdF	1	pString	Size Standard file name
STYP	1	cString	GeneMapper [®] Sample Type
SVER	1	pString	Data collection software version
SVER	3	pString	Data collection firmware version
SVER	4	pString	Sample File Format Version, containing the version of the sample file format used to write the file

Table 8 fsa File Tags from Applied Biosystems 3730/3730xl Data Collection Software v3.0 on the Applied Biosystems 3730/3730xl DNA Analyzer (*continued*)

Name	Number	ABIF Type	Description
Tmpr	1	long	Run temperature setting (degrees C)
TUBE	1	pString	Well ID
User	1	pString	Name of user who created the plate (optional)

SeqScape® Software v2.5 Tags

Optional tags are shown in italic text.

Table 9 abi File Tags from SeqScape® Software v2.5

Name	Number	ABIF Type	Description
AEPt	1	short	Stop point (orig)
AEPt	2	short	Stop point
ARTN	1	long	Analysis return code
ASPF	1	short	Basecaller adaptive processing success flag (legacy)
ASPt	1	short	Start point (orig)
ASPt	2	short	Start point
B1Pt	1	short	Peak 1 location (orig)
B1Pt	2	short	Peak 1 location
BCTS	1	pString	Time stamp for last successful basecalling analysis
DATA	9	short	Channel 1 analyzed data
DATA	10	short	Channel 2 analyzed data
DATA	11	short	Channel 3 analyzed data
DATA	12	short	Channel 4 analyzed data
<i>DATA</i>	<i>205</i>	<i>short</i>	<i>Channel 5 analyzed data (optional)</i>
Feat	1	user	Factura features
FTab	1	user	Factura features
FVoc	1	user	Factura features
PBAS	1	char	Basecalled sequence (edited)
PBAS	2	char	Basecalled sequence
PCON	1	char	Per-base quality values (edited)
PCON	2	char	Per-base quality values
PDMF	2	pString	Mobility file
phAR	1	float	Peak area ratio (equivalent to TRACE_PEAK_AREA_RATIO in phd1 file)
phCH	1	pString	Primer or terminator chemistry (equivalent to CHEM in phd1 file)
phDY	1	pString	Dye type (equivalent to DYE in phd1 file)
phQL	1	short	Maximum quality value (equivalent to QUALITY_LEVELS in phd1 file)
phTR	1	short	Read positions of first and last bases in trim region (along with pHTR below, equivalent to TRIM in phd1 file)
phTR	2	float	Trim probability threshold used (see above)
PLOC	1	short array	Base locations (edited)
PLOC	2	short array	Base locations
S/N%	1	short array	Signal level for each dye.

Table 9 abi File Tags from SeqScape® Software v2.5 (continued)

Name	Number	ABIF Type	Description
SPAC	1	float	Average peak spacing used in last analysis
SPAC	2	pString	Basecaller BCP/DLL
SPAC	3	float	Spacing
SVER	2	pString	Basecaller version

Sequencing Analysis Software v5.2 Tags

Optional tags are shown in italic text.

Table 10 abi File Tags from Sequencing Analysis Software v5.2

Name	Number	ABIF Type	Description
AEPt	1	short	Stop point (orig)
AEPt	2	short	Stop point
ARTN	1	long	Analysis return code
ASPF	1	short	Basecaller adaptive processing success flag (legacy)
ASPt	1	short	Start point (orig)
ASPt	2	short	Start point
B1Pt	1	short	Peak 1 location (orig)
B1Pt	2	short	Peak 1 location
BCTS	1	pString	Time stamp for last successful basecalling analysis
DATA	9	short	Channel 1 analyzed data
DATA	10	short	Channel 2 analyzed data
DATA	11	short	Channel 3 analyzed data
DATA	12	short	Channel 4 analyzed data
<i>DATA</i>	<i>205</i>	<i>short</i>	<i>Channel 5 analyzed data (optional)</i>
<i>Feat</i>	<i>1</i>	<i>user</i>	<i>Factura features (optional)</i>
<i>FTab</i>	<i>1</i>	<i>user</i>	<i>Factura features (optional)</i>
<i>FVoc</i>	<i>1</i>	<i>user</i>	<i>Factura features (optional)</i>
<i>NOIS</i>	<i>1</i>	<i>float array</i>	<i>The estimate of noise for each dye for a successfully analyzed file, only for the KB™ Basecaller. (optional)</i>
PBAS	1	char	Basecalled sequence (edited)
PBAS	2	char	Basecalled sequence
PCON	1	char	Per-base quality values (edited)
PCON	2	char	Per-base quality values
PDMF	2	pString	Mobility file
phAR	1	float	Peak area ratio (equivalent to TRACE_PEAK_AREA_RATIO in phd1 file)
phCH	1	pString	Primer or terminator chemistry (equivalent to CHEM in phd1 file)
phDY	1	pString	Dye type (equivalent to DYE in phd1 file)
phQL	1	short	Maximum quality value (equivalent to QUALITY_LEVELS in phd1 file)
phTR	1	short	Read positions of first and last bases in trim region (along with phTR below, equivalent to TRIM in phd1 file)
phTR	2	float	Trim probability threshold used (see above)
PLOC	1	short array	Base locations (edited)

Table 10 abi File Tags from Sequencing Analysis Software v5.2 (continued)

Name	Number	ABIF Type	Description
PLOC	2	short array	Base locations
RevC	1	short	Reverse complement flag
S/N%	1	short	Signal level
SPAC	1	float	Average peak spacing used in last analysis
SPAC	2	pString	Basecaller BCP/DLL
SPAC	3	float	Spacing
SVER	2	pString	Basecaller version

© Copyright 2006, Applied Biosystems. All rights reserved.

For Research Use Only. Not for use in diagnostic procedures.

Information in this document is subject to change without notice. Applied Biosystems assumes no responsibility for any errors that may appear in this document. In no event shall Applied Biosystems be liable for incidental, special, multiple, or consequential damages in connection with or arising from the use of this document.

NOTICE TO PURCHASER: DISCLAIMER OF LICENSE

Purchase of this software product alone does not imply any license under any process, instrument or other apparatus, system, composition, reagent or kit rights under patent claims owned or otherwise controlled by Applied Biosystems Corporation, either expressly or by estoppel.

TRADEMARKS:

Applied Biosystems, AB(Design), ABI PRISM, GeneMapper, GeneScan, and SeqScape are registered trademarks and Applied Biosystems and KB are trademarks of Applied Biosystems Corporation or its subsidiaries in the U.S. and/or certain other countries.

All other trademarks are the sole property of their respective owners.

Worldwide Sales and Support

Applied Biosystems vast distribution and service network, composed of highly trained support and applications personnel, reaches 150 countries on six continents. For sales office locations and technical support, please call our local office or refer to our Web site at www.appliedbiosystems.com.

Applied Biosystems is committed to providing the world's leading technology and information for life scientists. Applied Biosystems Corporation consists of the Applied Biosystems and Applied Genomics businesses.

Headquarters

850 Lincoln Centre Drive
Foster City, CA 94404 USA
Phone: +1 650.638.5800
Toll Free (In North America): +1 800.345.5224
Fax: +1 650.638.5884

07/2006