XMAP File Format Specification Sheet

Document Number: 30040

Document Revision: B

Legal Notices

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

Patents

Products of Bionano Genomics may be covered by one or more U.S. or foreign patents.

Trademarks

The Bionano Genomics logo and names of Bionano Genomics products or services are registered trademarks or trademarks owned by Bionano Genomics in the United States and certain other countries.

Bionano Genomics®, Irys®, IrysView®, IrysChip®, IrysPrep®, IrysSolve®, Saphyr™, Saphyr Chip™, and Bionano Access™ are trademarks of Bionano Genomics, Inc. All other trademarks are the sole property of their respective owners.

No license to use any trademarks of Bionano Genomics is given or implied. Users are not permitted to use these trademarks without the prior written consent of Bionano Genomics. The use of these trademarks or any other materials, except as permitted herein, is expressly prohibited and may be in violation of federal or other applicable laws.

XMAP v0.2 File Format Specification Sheet

This file format specification sheet details the file format specifications for the XMAP file, version 0.2. This file is backwards-compatible with XMAP v0.1.

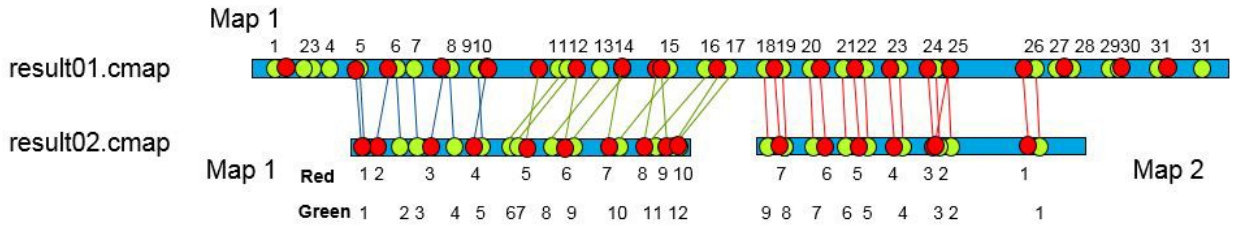
Introduction

The Bionano Genomics .xmap file is a cross-comparison between two maps.

The .xmap file reports the comparison derived from the alignment between an anchor .cmap file and a query .cmap file. The data line displays the map start and end coordinates and the locations of the labels on the map using a tab-delimited text based file.

The .xmap file presents the information in two sections: the XMAP information header, which describe the specific format of the data; and the map alignment information block, which contains the data rows. This file format specification sheet provides descriptions, with examples, of the XMAP header and map alignment information block format of the file.

When imported into Bionano Access™, the .xmap file is automatically filtered and ready for downstream analysis. XMAP files can be opened in Excel for easy readability or in any tab-delimited, text-based editor.



Format

The XMAP file contains the following sections:

- XMAP header
  - # XMAP File Version:
  - # Reference Maps From:
  - # Query Maps From:
  - #h
  - #f
- Alignment information block (each row as defined by the column headers in #h)
  - After the 3 IDs, is the first alignment of a reference map label to a query map label with orientation and confidence
  - Then the (pseudo)-CIGAR string displays in HitEnum, followed by query and reference length and label channel
  - The final string shows the alignment label site in the map and is repeated for all label sites indexed per label color channel

Header Specifications

Header rows are prefixed by the pound sign (#).

Table 1: Header Order

Header Line Tag	Header Line Description
# XMAP File Version:	Indicates the version of the XMAP file*
# Reference Maps From:	A string denoting the path to the corresponding _r.cmap*

Header Line Tag	Header Line Description
# Query Maps From:	A string denoting the path to the corresponding _q.cmap*
#h	Defines the columns for each data row**
#f	Defines the numerical data type for each data column**

Note: *Denotes the required header line tags for Bionano Access to read an XMAP file. Required header line tags must be present and must precede the Alignment Information Block to read an XMAP file. Header lines which are not required are optional and may be omitted. **Denotes the required header line tags for importing into Bionano Access.

Header Specification Details

The following tables provide the XMAP header's descriptions (including any specific formatting, limitations and requirements) and examples.

# XMAP File Version	
Header	# XMAP File Version:
Description	Indicates the version of the XMAP file.
Example	# XMAP File Version:<TAB>0.2

# Reference Maps From	
Header	# Reference Maps From:
Description	A string denoting the path to the corresponding reference map, which contains the reference or anchor data.
Example	# Reference Maps From:<TAB> ExampleXmap_r.cmap

# Query Maps From	
Header	# Query Maps From:
Description	A string denoting the path to the corresponding query map, which contains the query data.
Example	# Query Maps From:<TAB>ExampleXmap_q.cmap

#h		
Header	#h	
Description	Description of the required tab-separated columns in #h:	
	XmapEntryID	A unique line number for the data lines in the XMAP file. **Note:** For 2-color, the XmapEntryID will begin with the number 2.
	QryContigID	Map ID of query map (Contig ID from .cmap file for query)
	RefContigID	Map ID of the reference map from the .cmap reference file (the .cmap file may contain multiple reference maps). **Note:** RefContigIDs must be integers, but they need not be sequential.
	QryStartPos	Coordinates of the first aligned label on the query map (Start position of hit on query map)

#h		
	QryEndPos	Coordinates of the last aligned label on the query map (Stop position of hit on query map)
	RefStartPos	Coordinates of the first aligned label on the reference or anchor map
	RefEndPos	Coordinates of the last aligned label on the reference or anchor map
	Orientation	The relative orientation of the query map relative to the reference: forward (+) or reverse (-). The convention is that the reference is always positive orientation, so if the query aligns in reverse, it is shown as having negative (-) orientation. **Note:** For 2-color, the orientation will be the same.
	Confidence	Statistical Confidence of result: Negative Log10 of p-value of alignment (without Bonferroni Correction for multiple experiments). **Note:** For 2-color, the confidence number is the combined confidence of the alignment for both colors.
	HitEnum	Pseudo-CIGAR string representing matches (M), insertions (I), or deletions (D) of label sites with respect to the reference or anchor map. Count begins at the leftmost anchor label of that color. **Note:** When 2 or more anchor sites resolve into a single query site, only the rightmost anchor site is shown matched with the query site and the leftmost associated anchor sites are shown as deletions.
	QryLen	Length of query map from _q.cmap.
	RefLen	Length of reference map from _r.cmap.

	LabelChannel	Color channel of alignment from cmap files.
		• For 1-color data, LabelChannel is 1.
		• For 2-color data:
		  • Using -usecolor N, the LabelChannel is *N* (*N* = 1 or 2), and there is only one XMAP entry per alignment for the color channel specified by N.
		  • Without -usecolor N, LabelChannel is 1 or 2. In this case, there are two XMAP entries (two lines), one for each color channel.

#h		
	Alignment	Indices of the aligned site ID pairs. (When the query orientation is reversed ("-"), the query IDs are in descending order.) Count begins at the leftmost anchor label of that color. **Note:** When two sites in the reference align with the same site in the query, it is an indication that the two sites in the reference failed to resolve. Alignment provides a view of aligned pairs which would normally be ignored by HitEnum (CIGAR string).
Example	#h XmapEntryID<TAB>QryContigID<TAB> RefContigID <TAB>QryStartPos<TAB>QryEndPos<TAB>RefStartPos <TAB>RefEndPos<TAB>Orientation<TAB>Confidence <TAB>HitEnum <TAB>QryLen<TAB>RefLen<TAB>LabelChannel<TAB>Alignment	

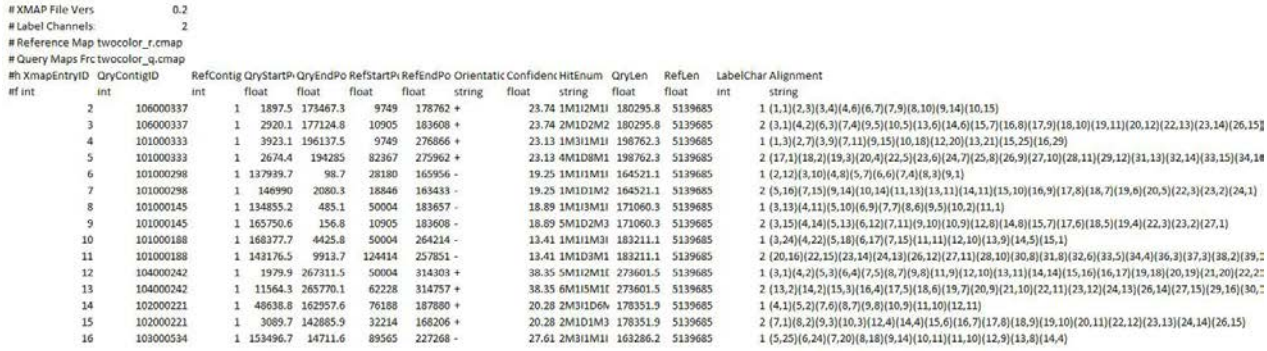
Note: Additional columns may be present but are not defined by XMAP Version 0.2.

#f	
Header	#f
Description	Defines the numerical data type for each data column.
Example	#f int<TAB>int<TAB>int<TAB>float<TAB>float<TAB> float<TAB>float<TAB>string<TAB>float<TAB>string

Alignment Information Block Specification

The data is grouped such that each data row represents an alignment between one reference or anchor map and one query contig/map.

Note: Depending on the parameters used during alignment, there may be more than one alignment for each reference and/or query map. Even for the same query and reference ID pair, different local alignments (alignments of the same region of the query with different regions of the reference) can be present.



# XMAP File Vers	0.2
# Label Channels:	2
# Reference Map	twocolor_r.cmap
# Query Maps Frc	twocolor_q.cmap

#h XmapEntryID	QryContigID	RefContig	QryStartPo	QryEndPo	RefStartPo	RefEndPo	Orientatic	Confidenc	HitEnum	QryLen	RefLen	LabelChar	Alignment
#f int	int	int	float	float	float	float	string	float	string	float	float	int	string
2	106000337	1	1897.5	173467.3	9749	178762	+	23.74	1M1I2M1I	180295.8	5139685		1 (1,1)(2,3)(3,4)(4,6)(6,7)(7,9)(8,10)(9,14)(10,15)
3	106000337	1	2920.1	177124.8	10905	183608	+	23.74	2M1D2M2	180295.8	5139685		2 (3,1)(4,2)(6,3)(7,4)(9,5)(10,5)(13,6)(14,6)(15,7)(16,8)(17,9)(18,10)(19,11)(20,12)(22,13)(23,14)(26,15)
4	101000333	1	3923.1	196137.5	9749	276866	+	23.13	1M3I1M1I	198762.3	5139685		1 (1,3)(2,7)(3,9)(7,11)(9,15)(10,18)(12,20)(13,21)(15,25)(16,29)
5	101000333	1	2674.4	194285	82367	275962	+	23.13	4M1D8M1	198762.3	5139685		2 (17,1)(18,2)(19,3)(20,4)(22,5)(23,6)(24,7)(25,8)(26,9)(27,10)(28,11)(29,12)(31,13)(32,14)(33,15)(34,1
6	101000298	1	137939.7	98.7	28180	165956	-	19.25	1M1I1M1I	164521.1	5139685		1 (2,12)(3,10)(4,8)(5,7)(6,6)(7,4)(8,3)(9,1)
7	101000298	1	146990	2080.3	18846	163433	-	19.25	1M1D1M2	164521.1	5139685		2 (5,16)(7,15)(9,14)(10,14)(11,13)(13,11)(14,11)(15,10)(16,9)(17,8)(18,7)(19,6)(20,5)(22,3)(23,2)(24,1)
8	101000145	1	134855.2	485.1	50004	183657	-	18.89	1M1I3M1I	171060.3	5139685		1 (3,13)(4,11)(5,10)(6,9)(7,7)(8,6)(9,5)(10,2)(11,1)
9	101000145	1	165750.6	156.8	10905	183608	-	18.89	5M1D2M3	171060.3	5139685		2 (3,15)(4,14)(5,13)(6,12)(7,11)(9,10)(10,9)(12,8)(14,8)(15,7)(17,6)(18,5)(19,4)(22,3)(23,2)(27,1)
10	101000188	1	168377.7	4425.8	50004	264214	-	13.41	1M1I1M3I	183211.1	5139685		1 (3,24)(4,22)(5,18)(6,17)(7,15)(11,11)(12,10)(13,9)(14,5)(15,1)
11	101000188	1	143176.5	9913.7	124414	257851	-	13.41	1M1D3M1	183211.1	5139685		2 (20,16)(22,15)(23,14)(24,13)(26,12)(27,11)(28,10)(30,8)(31,8)(32,6)(33,5)(34,4)(36,3)(37,3)(38,2)(39,
12	104000242	1	1979.9	267311.5	50004	314303	+	38.35	5M1I2M1I	273601.5	5139685		1 (3,1)(4,2)(5,3)(6,4)(7,5)(8,7)(9,8)(11,9)(12,10)(13,11)(14,14)(15,16)(16,17)(19,18)(20,19)(21,20)(22,
13	104000242	1	11564.3	265770.1	62228	314757	+	38.35	6M1I5M1I	273601.5	5139685		2 (13,2)(14,2)(15,3)(16,4)(17,5)(18,6)(19,7)(20,9)(21,10)(22,11)(23,12)(24,13)(26,14)(27,15)(29,16)(30,
14	102000221	1	48638.8	162957.6	76188	187880	+	20.28	2M3I1D6M	178351.9	5139685		1 (4,1)(5,2)(7,6)(8,7)(9,8)(10,9)(11,10)(12,11)
15	102000221	1	3089.7	142885.9	32214	168206	+	20.28	2M1D1M3	178351.9	5139685		2 (7,1)(8,2)(9,3)(10,3)(12,4)(14,4)(15,6)(16,7)(17,8)(18,9)(19,10)(20,11)(22,12)(23,13)(24,14)(26,15)
16	103000534	1	153496.7	14711.6	89565	227268	-	27.61	2M3I1M1I	163286.2	5139685		1 (5,25)(6,24)(7,20)(8,18)(9,14)(10,11)(11,10)(12,9)(13,8)(14,4)

Figure 1: Results map or .xmap