



Hybrid Scaffold Conflict Cut Status File Format Specification Sheet

Document Number: 30166

Document Revision: A

Legal Notices

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

This material is protected by United States Copyright Law and International Treaties. Unauthorized use of this material is prohibited. No part of the publication may be copied, reproduced, distributed, translated, reverse-engineered or transmitted in any form or by any media, or by any means, whether now known or unknown, without the express prior permission in writing from Bionano Genomics. Copying, under the law, includes translating into another language or format. The technical data contained herein is intended for ultimate destinations permitted by U.S. law. Diversion contrary to U. S. law prohibited. This publication represents the latest information available at the time of release. Due to continuous efforts to improve the product, technical changes may occur that are not reflected in this document. Bionano Genomics reserves the right to make changes in specifications and other information contained in this publication at any time and without prior notice. Please contact Bionano Genomics Customer Support for the latest information.

BIONANO GENOMICS DISCLAIMS ALL WARRANTIES WITH RESPECT TO THIS DOCUMENT, EXPRESSED OR IMPLIED, INCLUDING BUT NOT LIMITED TO THOSE OF MERCHANTABILITY OR FITNESS FOR A PARTICULAR PURPOSE. TO THE FULLEST EXTENT ALLOWED BY LAW, IN NO EVENT SHALL BIONANO GENOMICS BE LIABLE, WHETHER IN CONTRACT, TORT, WARRANTY, OR UNDER ANY STATUTE OR ON ANY OTHER BASIS FOR SPECIAL, INCIDENTAL, INDIRECT, PUNITIVE, MULTIPLE OR CONSEQUENTIAL DAMAGES IN CONNECTION WITH OR ARISING FROM THIS DOCUMENT, INCLUDING BUT NOT LIMITED TO THE USE THEREOF, WHETHER OR NOT FORESEEABLE AND WHETHER OR NOT BIONANO GENOMICS IS ADVISED OF THE POSSIBILITY OF SUCH DAMAGES.

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.

© Copyright 2017 Bionano Genomics, Inc. All rights reserved.

Conflict Cut Status File Format Specification Sheet

This file format specification sheet details the file format specifications for Conflict Cut Status File Version 0.1. It is highly recommended that conflicts are viewed and managed within Bionano Access. Bionano Access provides advanced visual features and an audit trail to simplify the management of automatic cuts and introduction of new cuts in both the maps and NGS contigs.

Introduction

In the Hybrid Scaffold pipeline, one important step is to identify and resolve conflicts. A conflict junction is the start position of a region where multiple data sources (e.g. NGS vs BNG) disagree, thus indicating possible assembly errors (see Figure 1). The Bionano Genomics conflict cut status file is a tab-delimited data file, which provides location information and the resolution strategy for each conflict junction detected in hybrid scaffold. The file can be opened in Excel for easy readability or in any tab-delimited, text-based editor.

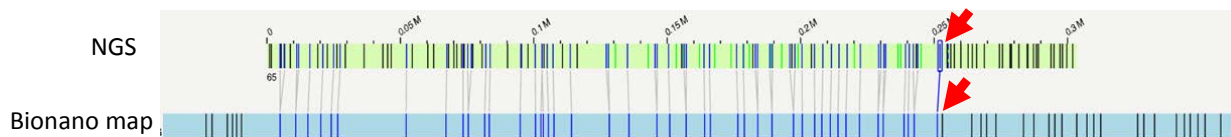


Figure 1. Example conflict between sequence and Bionano map. A significant number of unaligned labels outside the aligned region (left of the red arrows) indicates the presence of conflict between the two assemblies. Red arrows indicate the position of the conflict junction on the sequence and the genome map.

Format

The conflict cut status file contains the following sections:

- Header
 - #
 - #
- Conflict cuts information block:
 - First conflict
 - Next conflict [repeated for all conflicts]
 - Last conflict

Header Specifications

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

Header Line Tag	Header Line Description
#*	Defines the column name
#*	Defines the possible valid value for each column



Note: *Denotes the required header line tag to read a conflict cut status file.

Header Specification Details

The following tables provide the conflict cut status header's descriptions (including any specific formatting, limitations and requirement) and examples.

#	
Header	#
Description	Defines the columns for each data row in # rows:
xMapId	Each conflict is detected by an NGS-BNG alignment. This specifies the alignment id from which the conflicts is detected. Alignment Id is a non-negative number.
refQry	This column indicates whether an action is going to be performed on the NGS contig or BNG map. By convention, 'ref' refer to NGS contigs and 'qry' refers to BNG maps.
refId	The Id of the NGS contig where the conflict is detected. This can either be a positive Id number or '-1', which indicates this conflict is not associated with any NGS contig.
leftRefBkpt	A conflict is designated as "left" or "right" depending on whether the unaligned region is to the left or right of the conflict junction. A particular NGS-BNG alignment can have a conflict junction to the left or to the right or both. This column specifies the coordinate of a conflict junction on the NGS contig.
rightRefBkpt	The NGS coordinate of a conflict junction where the unaligned region is to the right of the site.
alignmentOrientation	'+' indicates the BNG map is aligned to the NGS contig with same orientation, '-' indicates opposite orientation.
ref_leftBkpt_toCut	This specifies whether to cut the NGS contig at the position specifies by 'leftRefBkpt'. The keyword 'cut' means cut the NGS contig and the keyword 'okay' means DO NOT cut the NGS contig.
ref_rightBkpt_toCut	This specifies whether to cut the NGS contig at the position specified by 'rightRefBkpt' (see previous column).
ref_toDiscard	The keyword 'exclude' means excluding the NGS contig specified by 'refId' from hybrid-scaffold (i.e. this NGS contig will not be used in any subsequent steps in hybrid scaffold). The keyword 'okay' means to keep the NGS contig in hybrid scaffolding.
refQry	Same as column 2, see above.
qryId	The Id of the BNG map from which the conflict is detected.
leftQryBkpt	Similar to 'leftRefBkpt', this specifies the coordinate of the conflict junction on the BNG map.

	rightQryBkpt	Similar to 'rightRefBkpt' but specifies the coordinate on BNG map instead.
	alignmentOrientation	See column 6.
	qry_leftBkpt_toCut	Similar to 'ref_right_Bkpt_toCut' but specifies the action performed for BNG map.
	qry_rightBkpt_toCut	Similar to 'ref_right_Bkpt_toCut'; see above.
	qry_toDiscard	Discard the BNG maps specified by 'qryId' (see ref_toDiscard').
Example	# xMapId<TAB>refQry<TAB> refId<TAB>leftRefBkpt<TAB>rightRefBkpt<TAB> alignmentOrientation<TAB>ref_leftBkpt_toCut<TAB>ref_rightBkpt_toCut<TAB> ref_toDiscard<TAB>refQry<TAB>qryId<TAB>leftQryBkpt<TAB>rightQryBkpt<TAB> alignmentOrientation<TAB>qry_leftBkpt_toCut<TAB>qry_rightBkpt_toCut<TAB> qry_toDiscard	

#	
Header	#
Description	Defines valid values allowed in each column. For each column, the set of valid values is separated by '/'. By convention, the last value in the set is a special value that indicates that this column is not relevant for this conflict. For example in the first column the values 'id/-1' indicates that this column can either be a valid xmap ID or '-1' which indicates this particular conflict is not associated with any NGS-BNG alignment (i.e. the user may have found this conflict using external data).
Example	# id/-1<TAB>ref<TAB>id/-1<TAB>position/-1<TAB>position/-1<TAB>+/-<TAB> okay/cut/-<TAB> okay/cut/-<TAB>okay/exclude/-<TAB>qry<TAB>id/-1<TAB> position/-1<TAB>position/-1<TAB>+/-<TAB>okay/cut/-<TAB>okay/cut/-<TAB> okay/exclude/-

Conflict cuts information block

- Conflict cuts information block
- First conflict
- Next conflict [repeated for all conflicts]
- Last conflict.

Example

# xMapId	refQry	refId	leftRefBkpt	rightRefBkpt	alignmentOrientation	ref_leftBkpt_toCut	ref_rightBkpt_toCut	ref_toDiscard	refQry	qryId	leftQryBkpt	rightQryBkpt	alignmentOrientation	qry_leftBkpt_toCut	qry_rightBkpt_toCut	qry_toDiscard
# id/-1	ref	id/-1	position/-1	position/-1	+/-	okay/cut/-	okay/cut/-	okay/exclude/-	qry	id/-1	position/-1	position/-1	+/-	okay/cut/-	okay/cut/-	okay/exclude/-
140	ref	96	-1	294721	-	okay	cut	okay	qry	5	-1	3541452	-	okay	okay	okay
660	ref	623	134613	-1	-	cut	okay	okay	qry	7	2790265	-1	-	okay	okay	okay
596	ref	548	-1	180307	-	okay	cut	okay	qry	27	-1	2853242	-	okay	okay	okay
326	ref	262	310534	-1	+	cut	okay	okay	qry	70	652954	-1	+	okay	okay	okay
105	ref	71	-1	329334	+	okay	cut	okay	qry	121	-1	906727	+	okay	okay	okay
89	ref	61	-1	188738	+	okay	cut	okay	qry	2	-1	2286421	+	okay	okay	okay
181	ref	130	469199	-1	+	cut	okay	okay	qry	107	1134675	-1	+	okay	okay	okay
532	ref	475	-1	322138	+	okay	cut	okay	qry	4	-1	5326556	+	okay	okay	okay

Figure 1. An Example of Conflict Cut Status File