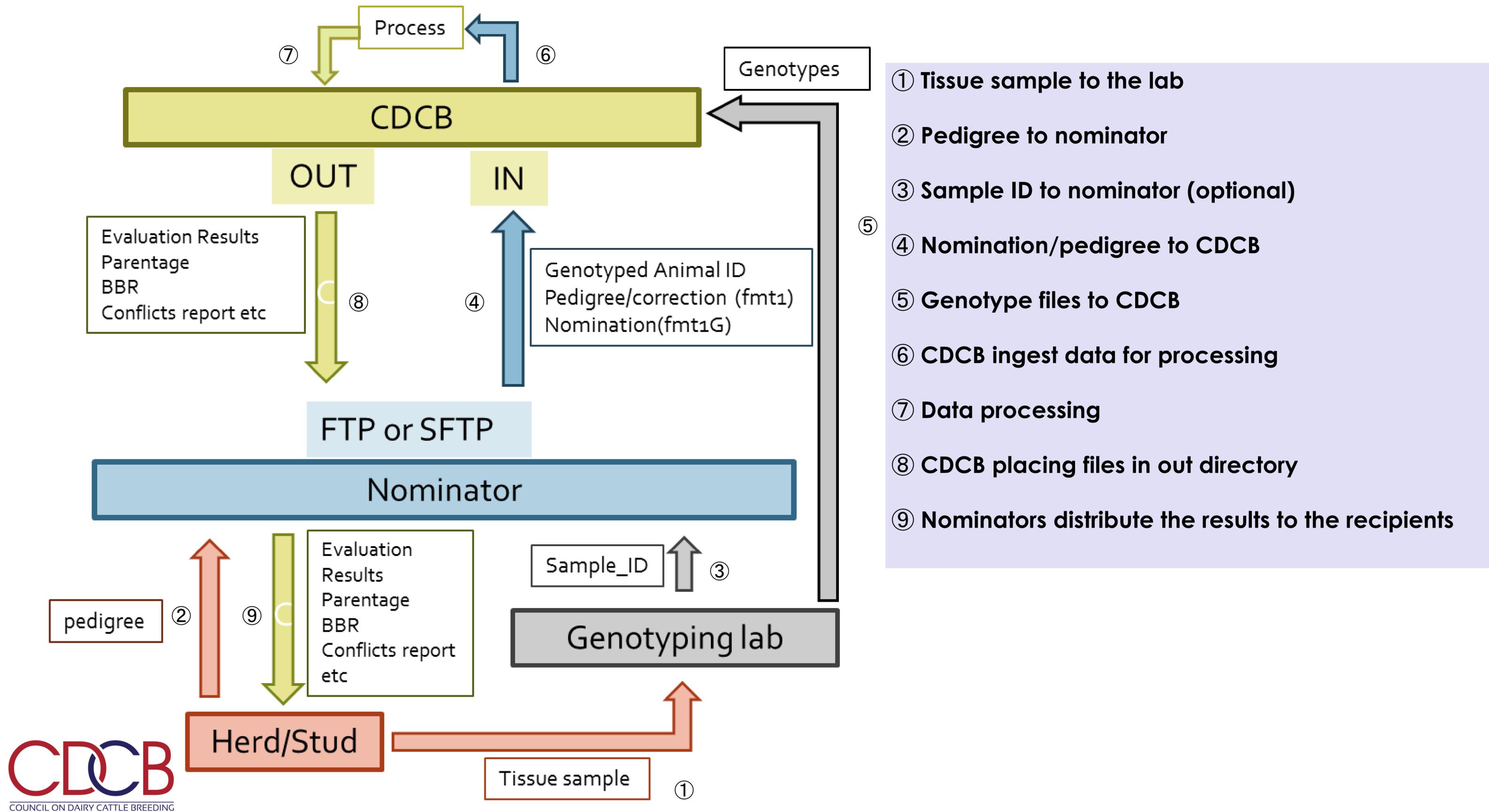# Files Exchanged and Conflict Resolution

5/27/2021

Kaori Tokuhisa, CDCB Genomic Data Analyst

- General files exchanged between CDCB and nominators and labs

- Conflict resolution

① Tissue sample to the lab

② Pedigree to nominator

③ Sample ID to nominator (optional)

④ Nomination/pedigree to CDCB

⑤ Genotype files to CDCB

⑥ CDCB ingest data for processing

⑦ Data processing

⑧ CDCB placing files in out directory

⑨ Nominators distribute the results to the recipients

https://redmine.uscdcb.com/projects/cdcb-customer-service/wiki/CDCB_general_files_distributed_to_nominators

Nomination (Step 4,6,7,8) – Before Genotype Loading!!

Input File (YYYYMMDD.1GX….)

- Nominate animals through CDCB Web query

- Nominate animals by submitting format1G file to CDCB SFTP "in" directory

0FBSUSA000068174286BSUSA000000198772BSUSA000068117173A1B2C3D4E5F6   20160219B20160312G013HR000000BELLA DEW SUGAR PIE ET   35051162N2P

- SampleID @54-70 (Optional)
- Record Type @88
- GroupID @130-137
- Genomic Fee code @139
- Herd code difference reason code @140 (if applicable)

Output Files (notify/format1E)

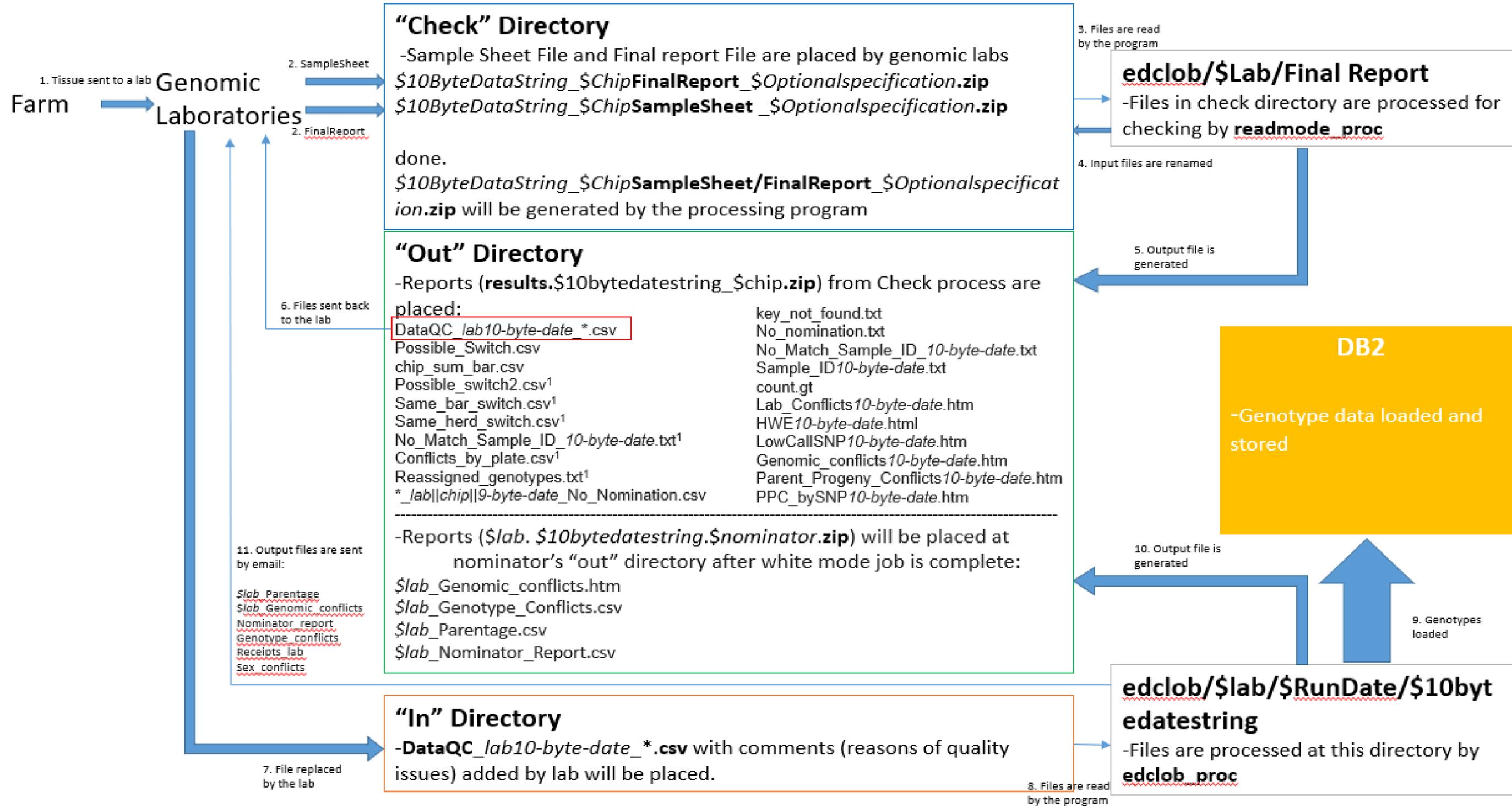**notify_VAL.YYYYMMDD.1GX –** Confirmation or Rejection of nomination and the reasons

notify_ERR.YYYYMMDD.1GX – Parentage and pedigree verification (not generated when there are no errors/changes)

YYYYMMDD.1EGX-Combination of the two above, in one line per animal. Not human friendly but good for computer

CDCB
COUNCIL ON DAIRY CATTLE BREEDING

Format1: https://redmine.uscdcb.com/projects/cdcb-customer-service/wiki/Format_1
Format1E: https://redmine.uscdcb.com/projects/cdcb-customer-service/wiki/Format_1E
Format1G: https://redmine.uscdcb.com/projects/cdcb-customer-service/wiki/Create_a_format1G
Error codes: https://queries.uscdcb.com/formats/geterr.cfm

Genotype file loading process (Step 4,5,6,7,8)

Farm

1. Tissue sent to a lab

Genomic Laboratories

2. SampleSheet

2. FinalReport

"Check" Directory
-Sample Sheet File and Final report File are placed by genomic labs
$10ByteDataString_$Chip**FinalReport**_$Optionalspecification.**zip**
$10ByteDataString_$Chip**SampleSheet** _$Optionalspecification.**zip**

done.
$10ByteDataString_$Chip**SampleSheet/FinalReport**_$Optionalspecification.**zip** will be generated by the processing program

3. Files are read by the program

edclob/$Lab/Final Report
-Files in check directory are processed for checking by **readmode_proc**

4. Input files are renamed

5. Output file is generated

"Out" Directory
-Reports (**results.**$10bytedatestring_$chip.**zip**) from Check process are placed:

DataQC_lab10-byte-date_*.csv	key_not_found.txt
Possible_Switch.csv	No_nomination.txt
chip_sum_bar.csv	No_Match_Sample_ID_10-byte-date.txt
Possible_switch2.csv[1]	Sample_ID10-byte-date.txt
Same_bar_switch.csv[1]	count.gt
Same_herd_switch.csv[1]	Lab_Conflicts10-byte-date.htm
No_Match_Sample_ID_10-byte-date.txt[1]	HWE10-byte-date.html
Conflicts_by_plate.csv[1]	LowCallSNP10-byte-date.htm
Reassigned_genotypes.txt[1]	Genomic_conflicts10-byte-date.htm
*_lab||chip||9-byte-date_No_Nomination.csv	Parent_Progeny_Conflicts10-byte-date.htm
	PPC_bySNP10-byte-date.htm

-Reports ($lab. $10bytedatestring.$nominator.**zip**) will be placed at nominator's "out" directory after white mode job is complete:
$lab_Genomic_conflicts.htm
$lab_Genotype_Conflicts.csv
$lab_Parentage.csv
$lab_Nominator_Report.csv

6. Files sent back to the lab

DB2

-Genotype data loaded and stored

10. Output file is generated

9. Genotypes loaded

11. Output files are sent by email:

$lab_Parentage
$lab_Genomic_conflicts
Nominator_report
Genotype_conflicts
Receipts_lab
Sex_conflicts

"In" Directory
-**DataQC_**lab10-byte-date_*.**csv** with comments (reasons of quality issues) added by lab will be placed.

7. File replaced by the lab

edclob/$lab/$RunDate/$10bytedatestring
-Files are processed at this directory by **edclob_proc**

8. Files are read by the program

CDCB
COUNCIL ON DAIRY CATTLE BREEDING

General Files Nominators receive

Frequency	Event	File Name	Description
Weekly	Evaluation	BRD_NOM_yyyymmdd.csv (xml) BRD_NOM_yyyymmdd_haplo_data.csv BRD_NOM.BBRdata.yyyymmdd.csv	Weekly evaluation,haplotype, and BBR result file for each breed (BRD), nominator(NOM), and date(yyyymmdd)
Monthly	Evaluation	NOM_YYMM.zip/NOM_YYMM.csv NOM_YYMM.zip/BRD_NOM_YYMM.xml NOM_YYMM.zip/NOM_yymm_haplo_data.csv NOM_Check_Fee_Code_YYMM.csv NOM_Report_Card.csv NOM_Supplemental.txt BRD_NOM_MGS_unlikely_YYMM.csv	-Monthly evaluation, BBR, result file for each breed (BRD), nominator(NOM), and month (YYMM) - a summary of fee codes submitted by the nominator, - a summary of the nominator's performance during the data submission period for the evaluation. - Supplemental information to report card - Suggested grand sires for animals with unlikely grand sires
	ID-based breed conflict	NOM_BRD_BBR90_BreedConflict_YYMM.csv	BBR > 90% in a different breed than the breed in the ID. Breed code must be updated for the genotype to become usable
Triannual	Evaluation	Evaluation results available online	Traditional+genomic evaluation available for all domestic males and females
Upon data submission	Quality Control	notify.yyyymmdd.1 NOM_LABCHIPYYYYMMDDXX_No_Nomination.csv NOM_Genomic_conflicts.htm NOM_Genomic_Errors.csv NOM_Parentage.csv NOM_Nominator_Report.csv NOM_PGS_unlikely.csv *Blue files above in LAB.YYYYMMDDXX.NOM.zip	-A report that indicates any error/notification of animals whose pedigree was processed -A file contains genotypes loaded without nomination -Genomic conflicts to be viewed on web browser -Genomic conflicts to be viewed /processed in server -Suggested parentage found by genomic information -A summary of the nominator's performance for the submitted batch -Suggested grandsires for animals with unlikely grandsires
Daily	Update (-o)	NOM_Parentage_yyyymmdd_TIMESTAMP.csv	Parentage suggested by genomic information and through program update

https://redmine.uscdcb.com/projects/cdcb-customer-service/wiki/CDCB_general_files_distributed_to_nominators

General Files Labs receive

Frequency	Event	File Name	Description
Monthly	Evaluation	LAB_Lab_Report_Card.csv	A summary of the nominator's performance during the data submission period for the evaluation.
Upon data submission	Check process	DataQC_lab10-byte-date_*.csv Possible_Switch.csv chip_sum_bar.csv Possible_switch2.csv Same_bar_switch.csv Same_herd_switch.csv Conflicts_by_plate.csv Reassigned_genotypes.txt NOM_lab||chip||9-byte-date_No_Nomination.csv key_not_found.txt No_nomination.txt No_Match_Sample_ID_10-byte-date.txt Sample_ID10-byte-date.txt count.gt Lab_Conflicts10-byte-date.htm HWE10-byte-date.htm Parent_Progeny_Conflicts10-byte-date.htm LowCallSNP10-byte-date.htm Genomic_conflicts10-byte-date.htm (Check_Errors_20200911.csv)	-Describes pass/fail, count, and description for results of QC process. -Shows possible switch at the lab where the genotyping was done -List chips with high error counts -Show possible switch at farms where the tissue sample was taken -Tells potential conflicts that could be elimited by switching the animal ID within the plate -Shows potential switch that happened at a farm -Shows conflict on each plate to investigate potential switch or shifting -generated when lab submitted genotype files to assign the genotype to new animal ID -contains 3 items (Sample ID, Animal ID, Problem With Sample -contains Animal ID, Sample ID, and requester with no pedigree **AND** no nomination on DB -contains requester ID, sample ID, and animal ID (as submitted). -contains any Sample IDs which were in the FinalReport file, but not in the SampleSheet. -This file shows association between the anima lID and sample. -contains number of GT submitted, new GT, GT reassigned and chip ID number -This file shows Chip based number of conflict -Statistics showing HWE to check homozygosity in autosome. -This file shows parent-progeny conflicts per SNP -indicates call rate(%) on each SNP - shows genomic conflict(s)
	Loading Process	Genomic_Errors.htm Low_call_genotypes.csv 8 files in blue above	-shows genomic conflicts -shows genotypes with low call rate

Reasons that a genomic evaluation was not released

- 1. The genotype was not usable due to a conflict or having low call rate.
- 2. The genotype became usable after the genotypes were extracted
- 3. The breed of evaluation is not among those we generate evaluations for
- 4. The animal's genotype conflicted with its imputed dam
- 5. The genomic fee code is N (No fee paid), or H (Historic)
- 6. The bull has semen marketed and it is not a traditional evaluation release (AI code=P/G/A/F/L/I) (https://www.naab-css.org/status-code-allowance)
- 7. The owner of the bull is not located in the US, so the evaluation is not public
- 8. The genotype is designated parentage verification only
- 9. The bull is foreign, over 15 mo of age (born 486 days before the 1st day of the month of the release) and no AI Service Fee has been paid.
- 10. The animal ID used for searching evaluation is incorrect or cross-referenced (primary ID should be used for searching)

https://redmine.uscdcb.com/projects/cdcb-customer-service/wiki/Reasons_that_a_genomic_evaluation_was_not_released

Conflict Resolution

- There are 2 processes where we report conflict

   1. When nomination is done or pedigree is updated

      -> notify files, format1E files contains error codes regarding the data validity

   2. When a genotype is processed

      -> $NOM_Genomic_Errors.csv and $NOM_Genotype_Conflicts.csv are included in a zip
      file named $LAB.yyyymmddXX.$NOM.zip to inform you about genomic errors

Notify and format1E files

notify_ERR file (after parentage + pedigree check)

Spec	Sex	Animal	Sire	Dam	Xref		
0	F						

Birth	Animal Name	GrpName	PI US RSN ProcDate	Src	Type	Ver	MBC	Reg	Stat	Disp
20201210			20210421	N	P	0	1			C

Code	Dsp	Reason	ID	Date	Herd	Source
5Be	N	Birth <> Master		20171210		N
3Gi	C	Bad Dam; brth conflict	HO			

Error Codes

notify_VAL file (after nomination check)

Spec	Sex	Animal	Sire	Dam	Xref		
0	F						

Birth	Animal Name	GrpName	PI US RSN ProcDate	Src	Type	Ver	MBC	Reg	Stat	Disp
20201210			7 0 20210421	N	G	5				N

Code	Dsp	Reason	ID	Date	Herd	Source
0Ri	N	Nomination acceptance				Y

Format1E file (Combination of the two above)

0FHO	20210312N20210421G51	21400063 10N0Ri	Y

Format1E specific

CDCB
COUNCIL ON DAIRY CATTLE BREEDING

Example of Pedigree Correction

```
Spec Sex       Animal               Sire                    Dam                    Xref
 0    F

 Birth            Animal Name          GrpName   PI US RSN ProcDate Src Type Ver MBC Reg Stat Disp
20201210                                                   20210421  N   P    0   1             C

Code Dsp Reason                       ID            Date         Herd      Source
5Be   N  Birth <> Master                         20171210                    N
3Gi   C  Bad Dam; brth conflict    HO
```

Example

Correct the birth date
• **Correction**: Create a <u>Format 1</u> record, following these instructions
   • set birth to correct birth
   • set pedigree verification code (@89) to 2
   • If the change is over 5 years multiple records will be needed so each record changes the birth date by less than 5 years. Only a breed source record can change a breed source pedigree. If the existing record is for an animal with lactations or progeny and the input record is not breed source, the birth date change can not be more than 90 days (400 days with verification). If the existing birth date is estimated and the input record is not breed source, the maximum change is 21 mo (638 days).

```
0FHO            HOUSA          HO                20181210N20210422P211    00000
0FHO            HOUSA          HO                20191210N20210422P211    00000
0FHO            HOUSA          HO                20201210N20210422P211    00000
```

Genomic Error files

NOM_Genomic_Errors.csv (1 error per row, Multiple rows for one animal)

Error description

Another animal ID that have the conflict with

Animal name of the animal in the left column

Source code

| Nominator | 18-digit Animal ID | Sample ID | Animal Name | N3, Sire conflict, |
| Nominator | 18-digit Animal ID | Sample ID | Animal Name | O3, Discovered non-pedigree parent-progeny relationship, |

Genomic Error Codes

Requester_ID,ID18,Sample_ID,Animal_Name,Error_code,Error,Other_ID,Other_Name,Source

NOM_Genotype_Conflicts.csv (1row for one animal)

Date on the batch

Another ID1 that have the conflict with Error Code1

Another ID1 that have the conflict with Error Code2

| Nominator | 18-digit Animal ID | Sample ID | Animal Name | ,20210415,O3, | N3, |

Genomic Error Code1

Genomic Error Code2

Requester,ID18,Sample_ID,Name,date,code1,ID1,code2,ID2,code3,ID3,code4,ID4,code5,ID5,code6,ID6

Example of Genomic Error Correction

`N3,Sire conflict,`
`O3,Discovered non-pedigree parent-progeny relationship,`

Example

N3

- **Full text of the error** : N3 - Sire Conflict
- **Meaning:** Genotypes of the animal and its pedigree sire have many Mendelian inheritance conflicts. In other words, genotype information indicates that the pedigree sire is not correct. Note: Such checks at CDCB are not restricted to ICAR's 200 parentage SNPs; CDCB checks all usable SNPs in common.
- **Correction:** When such an event occurs, the CDCB system may provide a discovered sire (see O3 - Discovered parent). This process is based on identifying bulls with few Mendelian inheritance conflicts with the animal. If, after consulting with the customer, the discovered sire is accepted, then pedigree information must be updated as indicated in Corrections to pedigree.

O3

- **Full text of the error** : O3 - Discovered non-pedigree parent-progeny relationship
- **Meaning** : Based on low Mendelian inheritance conflicts the genotypes appear to be parent and progeny. Indication of progeny or sire/dam is determined based on birth date and sex code from each animals existing pedigree information.
- **Correction** : If after consulting with the customer, a discovered sire/dam is agreed to, then pedigree information must be updated. The pedigree update can be performed by either accepting the discovered sire/dam in the genotype query[Fix_FMT1] or submitting a Format 1 with the discovered sire/dam.

Steps

1. Check with the owner to see if O3 suggests the true sire
2. If above is the case, check the birthdate of the animal and the sire
3. Correct the birthdate accordingly and accept the sire indicated in O3 by submitting format1

CDCB
COUNCIL ON DAIRY CATTLE BREEDING

https://redmine.uscdcb.com/projects/cdcb-customer-service/wiki/Genomic_error_codes

Using Genotype Query to Fix Conflict

Genotype Queries

Link	Query Description
●	Genotype Query: show (and fix) the usability/error status of received genotypes
●	Genotype Move/Swap APP : correct bad genotype to animal assignments
●	CDCB-Nomination_Q : ADD/REMOVE/UPDATE nomination and genotype information
●	History of Genotype(s)
●	Check FMT1 records
●	Get Dam : Suggests a dam for a genotyped animal based on group name, birth date and discovered MGS.
●	Check Dam : Check animal pedigree and error information, and dam pedigree/progeny/calving dates.
●	SAMPLE_ID look up tool for DNA labs
●	Get CDCB Fee from herdcode
●	Retrieve Parentage Validation Record
●	Get herd code and CDCB Fee for specified animal ID
●	Find Grandsire
●	Generate Genotype reports based on requester ID
●	Genomic 'How Do I ?' Documentation
●	Generate Lab reports based on LAB ID

Return to Query Form

Input (Animal_ID)	Eval Breed	Blend Code	MGS Status	PGS Status	Use Weekly	Pend. Ped.
▬▬▬▬▬	▬▬▬	S	MGS Info	PGS Info		

Animal	Sample ID	Barcode	Position	Chip	Fee	Use Ind.	Pend. GT	Conf.	Sire Status	Dam Status	Progeny Count	Tissue Source	Lab	Requester	Load Date	SNP Read Qty.
▬▬▬▬	▬▬▬	▬▬▬	▬▬▬	▬▬	2	N withdraw		⇑	2	4	0		▬	▬	2021-05-15	54705

Genotype Confirmations

Confirmation Type	Confirmed Genotype		
	Barcode	Position	Confirm ID
progeny	▬▬▬	▬▬	▬▬▬▬

Genotype Conflicts

Pedigree Changed	Conflict Type		Conflicting Genotype			
	Message	Code	Barcode	Position	Conflicting ID	USE
	Sire conflict	N3	▬▬	▬▬	▬▬▬	Y
	Discovered Sire	O3	▬▬	▬▬	▬▬▬	Y

Fix_FMT1

This button appears only when there is a conflict

Example of FIX_FMT1

FMT1 Records:

S p	S e x	ANIM ID17	Sire ID17	Dam ID17	Alias/Clone ID17	Birth (YYYYMMDD)	S r c	Modify Date (YYYYMMDD)	Rec Type	Ped_Veri Code	Rec Ver	Multi Birth Code	Reg Stat	Future Use	Name(30)	Herd/Group	X	Fee	HRC
current CDCB pedigree data																		REQ: ▬▬▬▬	
0	F	▬▬▬▬	▬▬▬▬	▬▬▬▬		20170503	B	20191106	P	2	1	1	HR	000000	▬▬▬▬	92200108	2		
genotype based suggestion																			
sugestions are based on genotypes *and* other data such as current sire/dam/animal pedigree birthdates, sample labeling, etc...																			
0	F	▬▬▬▬	▬▬▬▬	▬▬▬▬		20170503	A	20210518	P	2	1	1	HR	000000	▬▬▬▬				
User Input																		REQ: ▬▬▬▬	
0	F	▬▬▬▬	▬▬▬▬	▬▬▬▬	▬▬▬▬	20170503	A	20210518	P	2	1	1	HR	000000	▬▬▬▬	92200108		2 ⌄	⌄

Submit Changes

Before Monthly Cutoff...

- Please check if all animals have fee code assigned

- Please check if all the conflicts are resolved

Summary

- It is important to know the work flow, formats, inputs, outputs, tools, and rules

- Very informative documentation is available in wiki portal page

  https://redmine.uscdcb.com/projects/cdcb-customer-service/wiki

- CDCB improves system everyday, so please check announcement from CDCB for new changes and implementations

- If you have any questions or concerns, please do not hesitate to open a Redmine ticket any time (Please indicate your questions or requests clearly)

CDCB
COUNCIL ON DAIRY CATTLE BREEDING

Any Questions or Concerns?