CDCB Data Flow and File Exchange

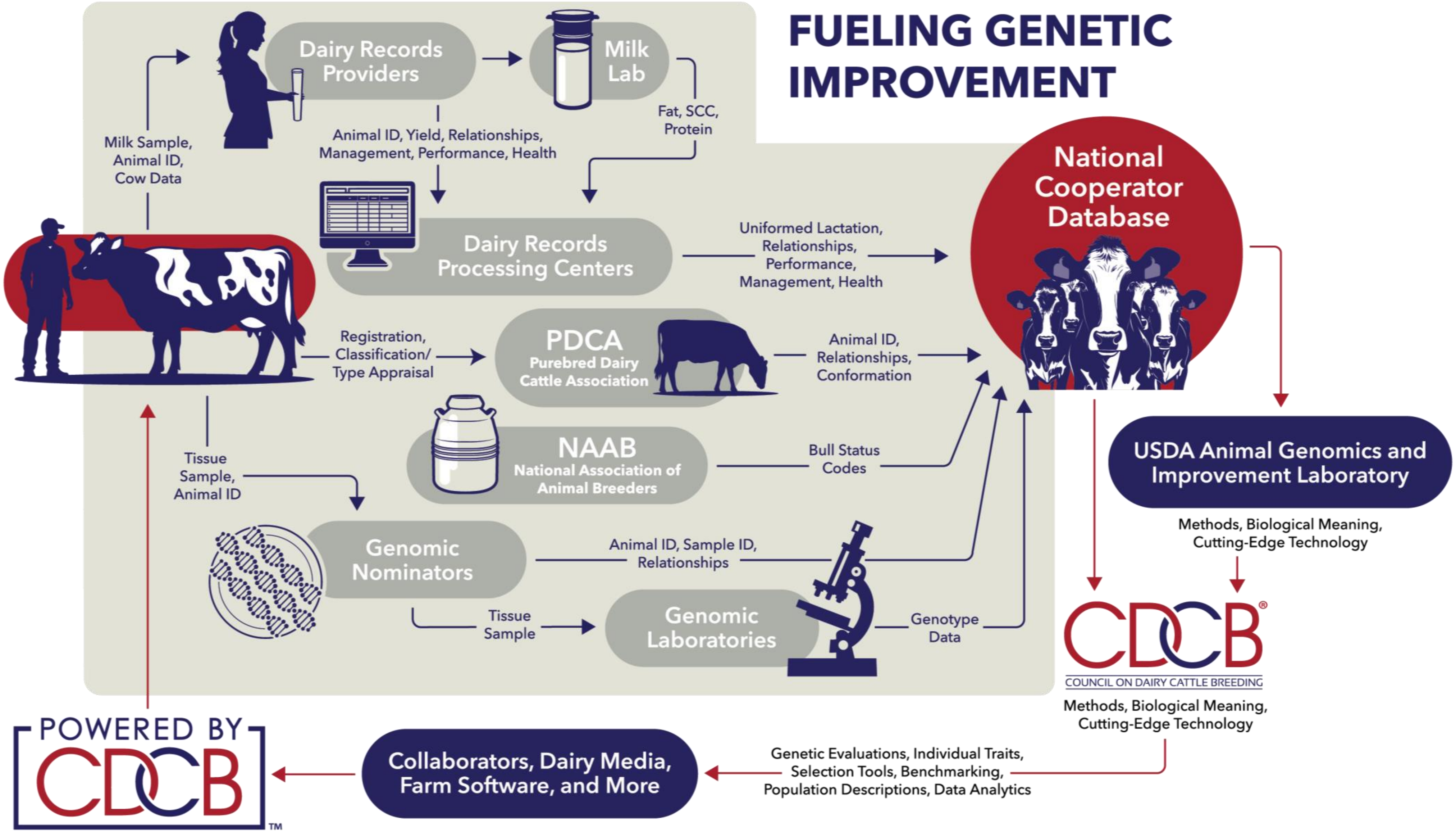
Hannah Templeton

Data Support Specialist

05/20/2026

NATIONAL COOPERATOR DATABASE

FUELING GENETIC IMPROVEMENT



Dairy Records Providers

Milk Lab

Milk Sample, Animal ID, Cow Data

Animal ID, Yield, Relationships, Management, Performance, Health

Fat, SCC, Protein

Dairy Records Processing Centers

Uniformed Lactation, Relationships, Performance, Management, Health

National Cooperator Database

Registration, Classification/ Type Appraisal

PDCA
Purebred Dairy Cattle Association

Animal ID, Relationships, Conformation

NAAB
National Association of Animal Breeders

Bull Status Codes

USDA Animal Genomics and Improvement Laboratory

Methods, Biological Meaning, Cutting-Edge Technology

Tissue Sample, Animal ID

Genomic Nominators

Animal ID, Sample ID, Relationships

Tissue Sample

Genomic Laboratories

Genotype Data

CDCB
COUNCIL ON DAIRY CATTLE BREEDING

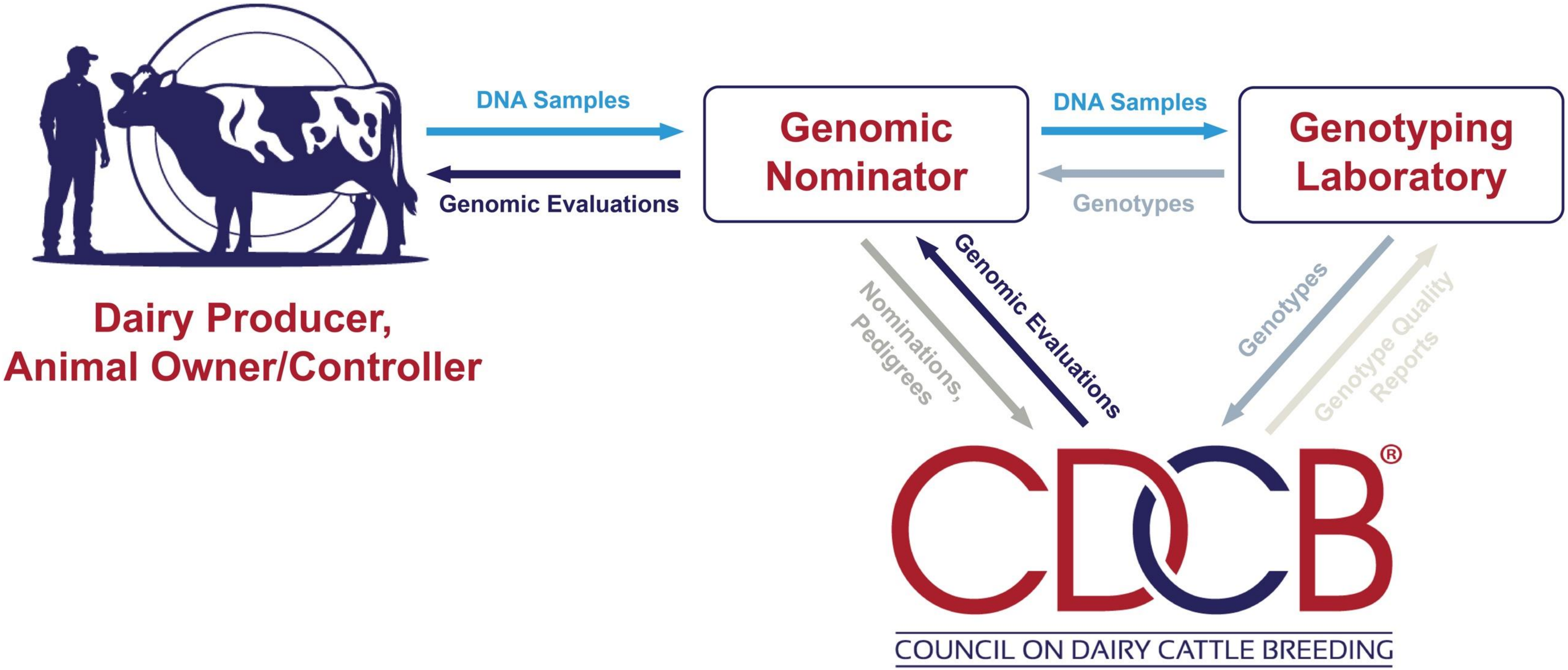
Methods, Biological Meaning, Cutting-Edge Technology

Genetic Evaluations, Individual Traits, Selection Tools, Benchmarking, Population Descriptions, Data Analytics

Collaborators, Dairy Media, Farm Software, and More

POWERED BY CDCB™

How to Request a Genomic Evaluation



DNA Samples

Genomic Evaluations

Genomic Nominator

DNA Samples

Genotypes

Genotyping Laboratory

Dairy Producer, Animal Owner/Controller

Nominations, Pedigrees

Genomic Evaluations

Genotypes

Genotype Quality Reports

CDCB®

COUNCIL ON DAIRY CATTLE BREEDING

Data Submission

Nominators

Nomination + pedigree

- Submission of format1 (G,P) through SFTP
- Submission through WebConnect

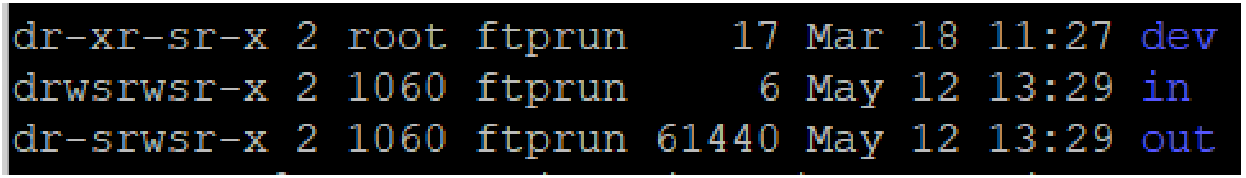
Labs

Genotype Files (Sample Sheet + Final Report)

- Submission through SFTP ***ONLY***

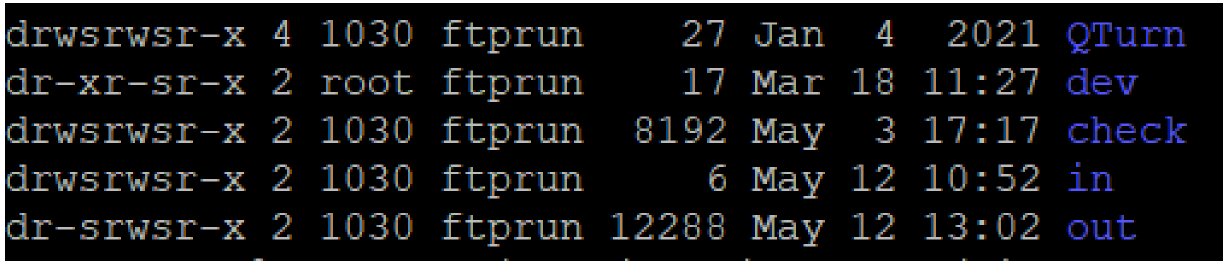
Secure File Transfer Protocol (SFTP)

Nominator SFTP



```
dr-xr-sr-x 2 root ftprun    17 Mar 18 11:27 dev
drwsrwsr-x 2 1060 ftprun     6 May 12 13:29 in
dr-srwsr-x 2 1060 ftprun 61440 May 12 13:29 out
```

Lab SFTP



```
drwsrwsr-x 4 1030 ftprun    27 Jan  4  2021 QTurn
dr-xr-sr-x 2 root ftprun    17 Mar 18 11:27 dev
drwsrwsr-x 2 1030 ftprun  8192 May  3 17:17 check
drwsrwsr-x 2 1030 ftprun     6 May 12 10:52 in
dr-srwsr-x 2 1030 ftprun 12288 May 12 13:02 out
```

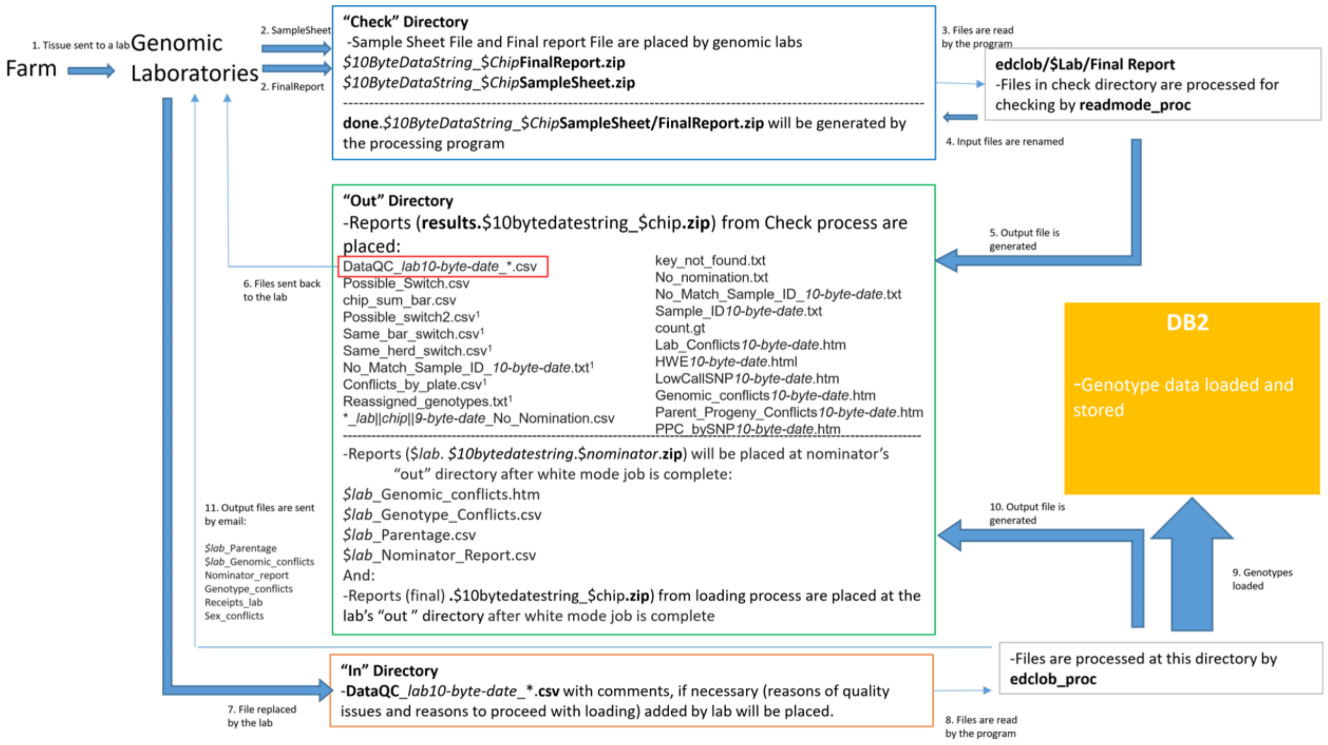
- Data exchanges via files happens in SFTP
- "in" folder for incoming files and "out" folder for outgoing files
- Lab gets additional folder called "check"
    - Genotypes can be loaded from SFTPs only
- Files get picked up by our automation if there is no pending job prior to your submission
- Some files in "out" directory are zipped after 24 hours to save storage (20240730.zip)

SFTP vs WebConenct

	SFTP	WebConnect
Ease of use	Required to follow format1 format	Graphical User Interface, user friendly
Number of records + submission	Batch submissions allowed	Limitation in number of submissions (Data exchange for batch is available for some data)
Record type flexibility	Most record types accepted (P, X, C, R, etc.)	Some record types not supported (such as C and D)
System	Some infrastructure might need to be built, but automation is possible	No system required, manual work involved

Genotype Submission

- Labs submit Sample Sheet files (contains sample info), and Final Report files (contains actual genotypes) in zipped files to “Check” directory for QC.
- Once QC is done, many QC report files are placed in “out” directory.
- **Data QC** file contains pass/fail from QC check
  - Put a comment (if there are any failed metrics) after investigating the issue and still want to proceed
  - No comments are required for submissio that pass QC check
- Placing **Data QC** file in “in” directory triggers the system to proceed with loading.



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

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

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

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

DataQC_*lab10-byte-date_**.csv
Possible_Switch.csv
chip_sum_bar.csv
Possible_switch2.csv[1]
Same_bar_switch.csv[1]
Same_herd_switch.csv[1]
No_Match_Sample_ID_*10-byte-date*.txt[1]
Conflicts_by_plate.csv[1]
Reassigned_genotypes.txt[1]
*_*lab||chip||9-byte-date*_No_Nomination.csv

key_not_found.txt
No_nomination.txt
No_Match_Sample_ID_*10-byte-date*.txt
Sample_ID*10-byte-date*.txt
count.gt
Lab_Conflicts*10-byte-date*.htm
HWE*10-byte-date*.html
LowCallSNP*10-byte-date*.htm
Genomic_conflicts*10-byte-date*.htm
Parent_Progeny_Conflicts*10-byte-date*.htm
PPC_bySNP*10-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
And:
-Reports (final) **.**$10bytedatestring_$chip**.zip**) from loading process are placed at the lab’s “out ” directory after white mode job is complete

11. Output files are sent by email:

*$lab*_Parentage
*$lab*_Genomic_conflicts
Nominator_report
Genotype_conflicts
Receipts_lab
Sex_conflicts

DB2

-Genotype data loaded and stored

“In” Directory
-**DataQC_***lab10-byte-date_****.csv** with comments, if necessary (reasons of quality issues and reasons to proceed with loading) added by lab will be placed.

-Files are processed at this directory by **edclob_proc**

General Files Distributed for Labs

- Zip files generated from **check** process:
  results. YYYYMMDDXX_ZZZ.zip

- Zip file generated from **write mode** process:
  final.results. YYYMMDDXX_ZZZ.zip

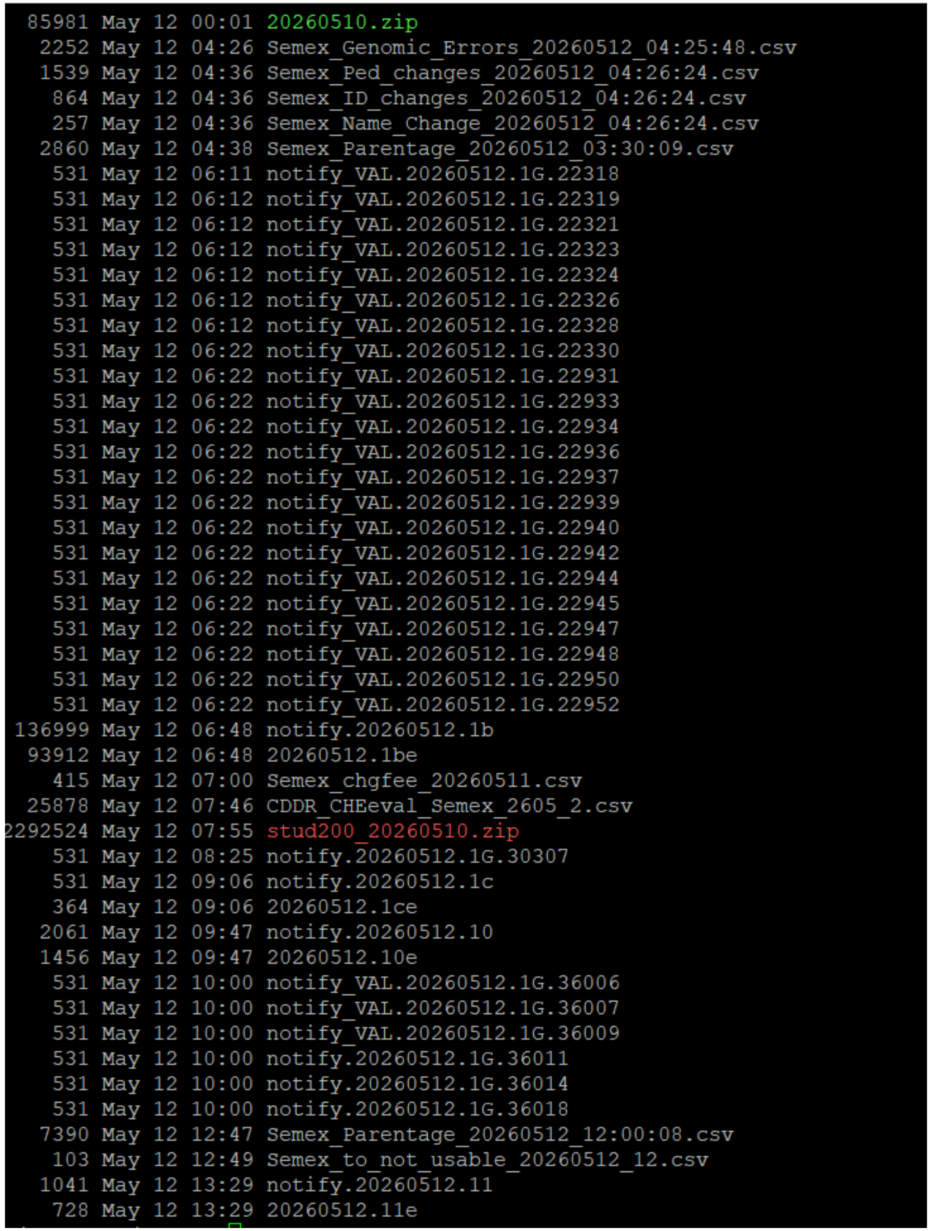
YYYY = Year (2024)
MM= Month (07)
DD = Day (30)
XX = Number of submission
ZZZ = Chip type (ex. 50K for Illumina BovineSNP50 BeadChip. V1)

- DataQC_lab10-byte-date_*.csv
- Possible_Switch.csv
- chip_sum_bar.csv
- Conflicts_by_plate.csv
- Reassigned_genotypes.txt
- 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.html
- Parent_Progeny_Conflicts10-byte-date.htm
- PPC_bySNP10-byte-date.htm
- LowCallSNP10-byte-date.htm
- Genomic_conflicts10-byte-date.htm (Check_Errors_20200911.csv)
- GE1k_progeny.csv
- Existing_genotype_negative_key.csv
- Excessive_Homozygous_LABYYYYMMDD.csv (upon loading)

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

General Files Distributed for Nominators

- Files are generated on 5 occasions: weekly, monthly, triannual, upon data submission, and daily processing (-o job)
- Files include information like: evaluation, notifications, QC, and updates
- Notify files might be the file that you check most frequently

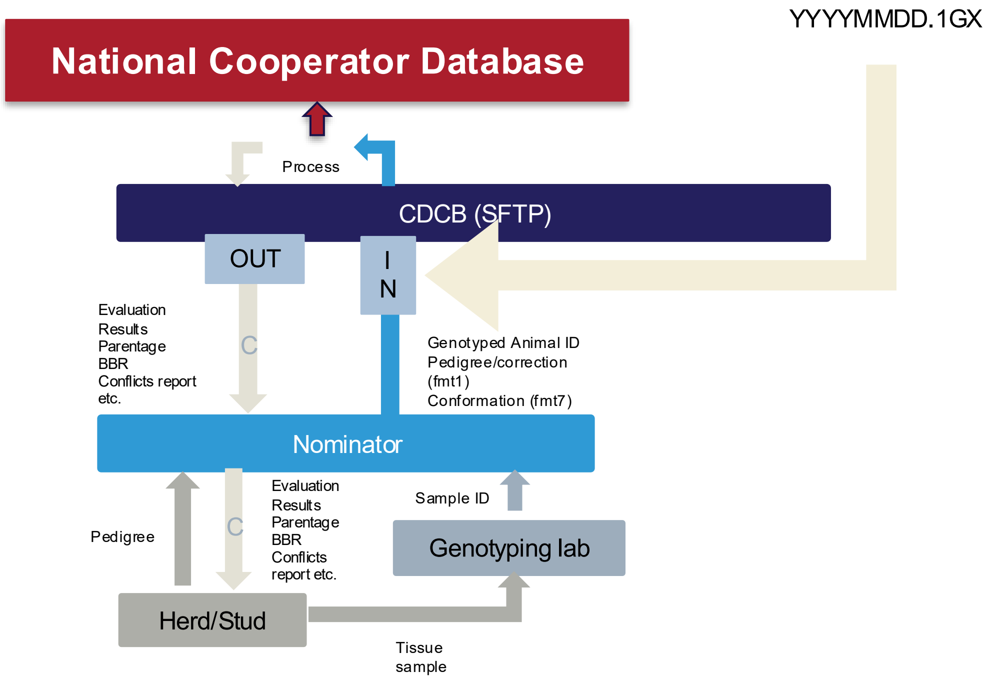


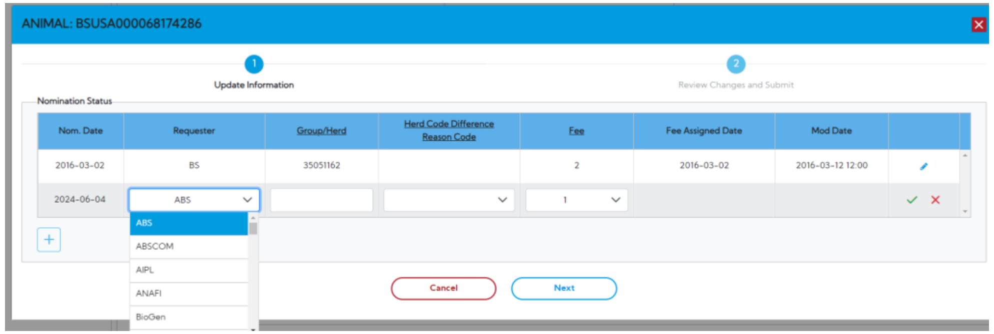
```
  85981 May 12 00:01 20260510.zip
   2252 May 12 04:26 Semex_Genomic_Errors_20260512_04:25:48.csv
   1539 May 12 04:36 Semex_Ped_changes_20260512_04:26:24.csv
    864 May 12 04:36 Semex_ID_changes_20260512_04:26:24.csv
    257 May 12 04:36 Semex_Name_Change_20260512_04:26:24.csv
   2860 May 12 04:38 Semex_Parentage_20260512_03:30:09.csv
    531 May 12 06:11 notify_VAL.20260512.1G.22318
    531 May 12 06:12 notify_VAL.20260512.1G.22319
    531 May 12 06:12 notify_VAL.20260512.1G.22321
    531 May 12 06:12 notify_VAL.20260512.1G.22323
    531 May 12 06:12 notify_VAL.20260512.1G.22324
    531 May 12 06:12 notify_VAL.20260512.1G.22326
    531 May 12 06:12 notify_VAL.20260512.1G.22328
    531 May 12 06:22 notify_VAL.20260512.1G.22330
    531 May 12 06:22 notify_VAL.20260512.1G.22931
    531 May 12 06:22 notify_VAL.20260512.1G.22933
    531 May 12 06:22 notify_VAL.20260512.1G.22934
    531 May 12 06:22 notify_VAL.20260512.1G.22936
    531 May 12 06:22 notify_VAL.20260512.1G.22937
    531 May 12 06:22 notify_VAL.20260512.1G.22939
    531 May 12 06:22 notify_VAL.20260512.1G.22940
    531 May 12 06:22 notify_VAL.20260512.1G.22942
    531 May 12 06:22 notify_VAL.20260512.1G.22944
    531 May 12 06:22 notify_VAL.20260512.1G.22945
    531 May 12 06:22 notify_VAL.20260512.1G.22947
    531 May 12 06:22 notify_VAL.20260512.1G.22948
    531 May 12 06:22 notify_VAL.20260512.1G.22950
    531 May 12 06:22 notify_VAL.20260512.1G.22952
 136999 May 12 06:48 notify.20260512.1b
  93912 May 12 06:48 20260512.1be
    415 May 12 07:00 Semex_chgfee_20260511.csv
  25878 May 12 07:46 CDDR_CHEeval_Semex_2605_2.csv
2292524 May 12 07:55 stud200_20260510.zip
    531 May 12 08:25 notify.20260512.1G.30307
    531 May 12 09:06 notify.20260512.1c
    364 May 12 09:06 20260512.1ce
   2061 May 12 09:47 notify.20260512.10
   1456 May 12 09:47 20260512.10e
    531 May 12 10:00 notify_VAL.20260512.1G.36006
    531 May 12 10:00 notify_VAL.20260512.1G.36007
    531 May 12 10:00 notify_VAL.20260512.1G.36009
    531 May 12 10:00 notify.20260512.1G.36011
    531 May 12 10:00 notify.20260512.1G.36014
    531 May 12 10:00 notify.20260512.1G.36018
   7390 May 12 12:47 Semex_Parentage_20260512_12:00:08.csv
    103 May 12 12:49 Semex_to_not_usable_20260512_12.csv
   1041 May 12 13:29 notify.20260512.11
    728 May 12 13:29 20260512.11e
```

0FHOUSA000011111111HOUSA000000222222HOUSA000033333333A1B2C3D4E5F6 20160219B20160312G013HR000000NAMEOFANIMAL ET 35051162N2

Nomination

- Nomination is a process of providing pedigree, assigning a fee code, and indicating what service you would like to receive
- Notify file will report any nomination related errors
- Nominations can be made on WebConnect through Bulk Nomination under the Special Section tab or under the Genotype Tab
- Nomination should be done before the genotype submission




ANIMAL: BSUSA000068174286

Nomination Status

Nom. Date	Requester	Group/Herd	Herd Code Difference Reason Code	Fee	Fee Assigned Date	Mod Date
2016-03-02	BS	35051162		2	2016-03-02	2016-03-12 12:00
2024-06-04	ABS			1		

https://redmine.uscdcb.com/projects/cdcb-customer-service/wiki/Create_a_format1G

https://redmine.uscdcb.com/projects/cdcb-customer-service/wiki/Format_1

Final Step

Please make sure the nomination/pedigree was processed successfully by looking at notify files in your **out** directory!

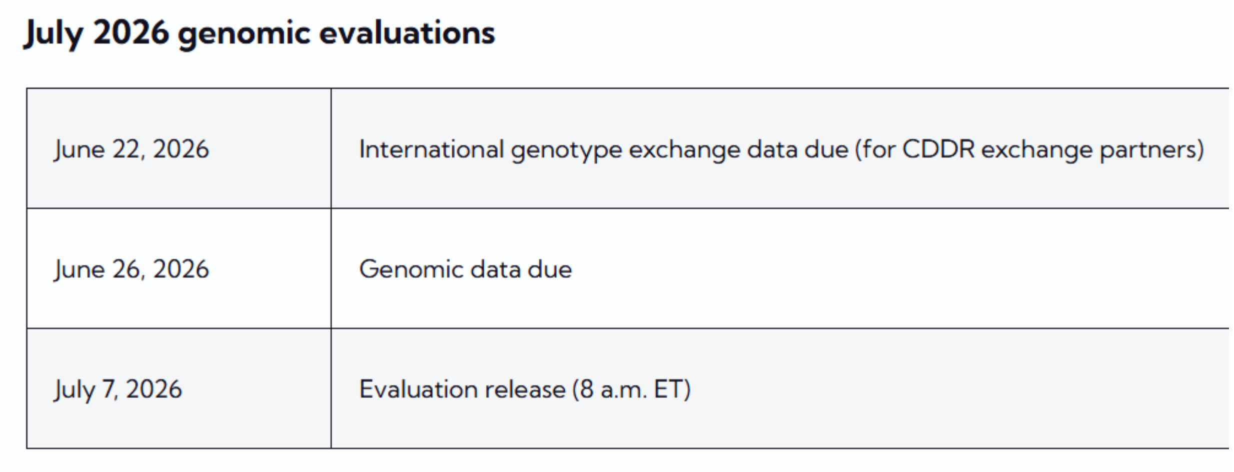
- **notify_VAL.YYYYMMDD.1GX**

Notify whether the animal was nominated successfully or not regardless of existence of errors.

- **notify_ERR.YYYYMMDD.1[G]X**

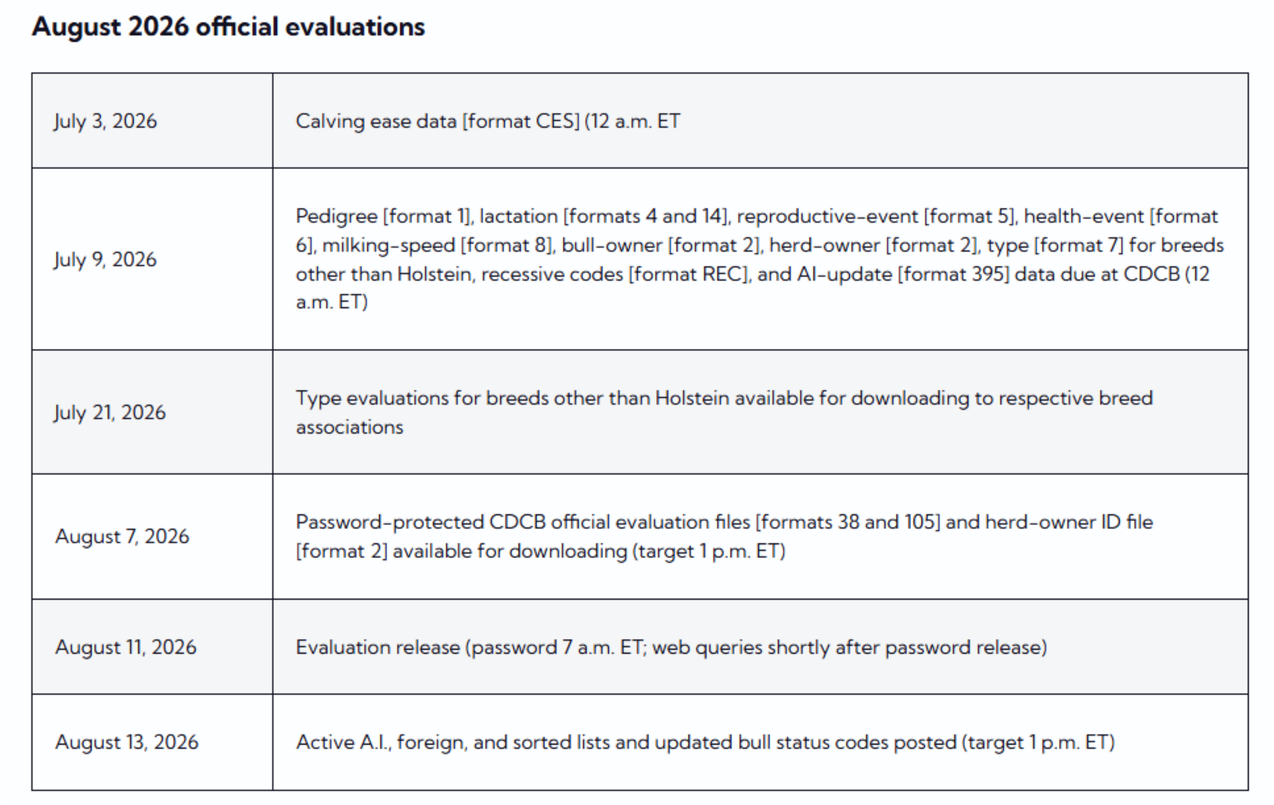
Notify only if error/change needs to be informed.

Data Submission Deadlines



July 2026 genomic evaluations

June 22, 2026	International genotype exchange data due (for CDDR exchange partners)
June 26, 2026	Genomic data due
July 7, 2026	Evaluation release (8 a.m. ET)



August 2026 official evaluations

July 3, 2026	Calving ease data [format CES] (12 a.m. ET
July 9, 2026	Pedigree [format 1], lactation [formats 4 and 14], reproductive-event [format 5], health-event [format 6], milking-speed [format 8], bull-owner [format 2], herd-owner [format 2], type [format 7] for breeds other than Holstein, recessive codes [format REC], and AI-update [format 395] data due at CDCB (12 a.m. ET)
July 21, 2026	Type evaluations for breeds other than Holstein available for downloading to respective breed associations
August 7, 2026	Password-protected CDCB official evaluation files [formats 38 and 105] and herd-owner ID file [format 2] available for downloading (target 1 p.m. ET)
August 11, 2026	Evaluation release (password 7 a.m. ET; web queries shortly after password release)
August 13, 2026	Active A.I., foreign, and sorted lists and updated bull status codes posted (target 1 p.m. ET)

https://uscdcb.com/genetic-evaluation-schedule/2026/

Summary

- Due to the variety and complexity of the data we exchange, CDCB outputs many kinds of files with different information (confirmation, error/conflicts, results, notification, etc.)
- Utilizing CDCB documentation, as well as your own internal documentation, should help with understanding these files

COUNCIL ON DAIRY CATTLE BREEDING

Thank you!