

ICTV 36 Data Load - 2020 (loaded in 2021)

Source	Location
change.xlsx	box:/ictvonline/taxonomy/ICTV_update/2020_updates ICTV MSL Release 36 2020 Changes.0.xlsx
db load code	https://github.com/rusalkaguy/ICTVonlineDbLoad \\genome-ws-02\ICTVonlineDbLoad\
email	"EC 52, Online, October 2020; Email ratification March 2021 (MSL #36)"
dev.ictvonline.org	RDC ec2-54-89-205-80.compute-1.amazonaws.com <ul style="list-style-type: none">• elliot copied .bak file from production: d:/MSSQL/Backup/ICTVonline35_042721.zip• Curtis restored that backup into ICTVonline36

Curtis,

The big day is here! The spreadsheet of ICTV taxon changes approved last summer and ratified this March is now ready for loading into the database. A few important items:

1. The data is located in the Box Share [ictvonline:taxonomy:ICTV_update:2020_updates](#)
2. The spreadsheet name is ICTV MSL Release 36 2020 Changes.0.xlsx
3. The release information is: "EC 52, Online, October 2020; Email ratification March 2021 (MSL #36)"
4. For this release, all type species have been abolished. There should be none specified in the excel sheet, and every species designated as a type species in MSL#35 should have that designation removed for MSL #36.
5. You can do your work on [dev.ictvonline.org](#). This is now an AWS VM that is equivalent to the production server. Things should be much faster. Your login is in Keeper under the AWS share, ICTVdev curtish. The RDP connection info is there as well.

Thanks, Elliot

0. Load MSL update data

- RDC connect, mounting local drive (E:\) that has the git clone of our SQL scripts
 - Create load_next_msl Table
 - rename existing load_next_msl table to add suffix for the MSL it was used for
 - \\tsclient\E\ICTVonlineDbLoad\0.b1. rename old load_next_msl.sql
 - change _## values to _35
 - run
 - prepare .xls file
 - save "excel_files\ICTV MSL Release 36 2020 Changes.1.xlsx" as "excel_files\ICTV MSL Release 36 2020 Changes.1.col_mapped.xls"
 - add column A and rows 1:2
 - copy in headers and col A filename macro from "col_mapped.xls" to edit column headers to match table (because you'll likely reload several times),
 - add SQL generation lines (copy from "col_mapped.xls" and adjust number of columns
 - 0.b2.create_table-load_next_msl-delta.sql
 - update MSL numbrer in CONSTRAINT [DF_load_next_msl__msl_release_num]
 - then edit for new/missing/columns in this MSL's sheet
 - run to create table
 - load MSL
 - Skip: MSSQL Import Data : xls driver not correctly installed.
- ERROR: The 'Microsoft.ACE.OLEDB.12.0' provider is not registered on the local machine
- copy SQL from Excel into "1.a.load_next_msl-delta-insert_values.sql"
 - run

0. Other schema Updates

- 0.b3.alter_schema.sql - no changes this MSL
- (other 0.b#.alter*.sql ?
- 0.c.insert_msl_into_taxonomy_toc.sql

- edit: msl='36'
- edit: root_name='2020'
- run
- **0.m.drop_obsolete_tables.sql** - no changes this MSL

1. Cleanup Data

- **1.0 load progress query.sql**
 - check that #rows in load_next_msl = # of data rows in source excel file: **3882**
- **Backup DB** to D:\MSSQL\Backup\ICTVonline36\ICTVonline36_20210430_0202_load_next_msl_filled.bak
- **"1.b.load_next_msl-convert_blank_to_null.sql"**
 - update to match current columns using "2.a.load_next_msl - data cleans.sql.xlsx"
 - run: fixed many values
- **"2.a.load_next_msl - data cleans.sql"**
 - update to match current columns using "2.a.load_next_msl - data cleans.sql.xlsx"
 - begin transaction
 - leading/training spaces: N=many
 - curly quotes: N=0
 - remove quotes around values: N=0
 - misc other errors at end: N=0
 - commit transaction
- **"2.a2.load_next_msl - data cleaning CSVs.sql"**
 - hand fix 2
 - Sort=42892 exemplarAccessions='L: MN567049 and MN567050; S:MN567048'
 - Sort=12246 exemplarAccessions='MK430076 (DNA-A), MK430077 (DNA-B), MK430078 (DNA-C)'
 - run standard "change to CSV" fixes
 - commit transaction
- **2.b map actions.sql**
 - add 'promote%'
 - remove isType logic
 - fix identical name check logic

2. Link to existing Taxonomy

- **Backup DB** to D:\MSSQL\Backup\ICTVonline36\ICTVonline36_20210430_1216_pre_alter_src_lineage_col.bak
- **2.c.load_next_msl-set-prev_tax_id.sql**
 - begin transaction
 - test, then update: rpt='Map _src_taxon_name to last MSL'; N=447
 - QC - unmapped rows: ERRORS=10
 - had to go back and correct load_next_msl_src_lineage which repeated srcRealm 3 extra times in the middle(!)
 - <https://github.com/rusalkaguy/ICTVonlineDbLoad/commit/e34bd6712078912abaabd0d531f8a3edb6b4ab27>
 - correct typos
 - correct srcGenus='unassigned'; change to NULL
 - Some renames correct errors in this same MSL. For those, we will combine records, using a ";" to separate proposal filenames, and mark the obsolete records with isWrong=primary.sort

primary sort	primary proposal	primary change	secondary sort	secondary proposal	2ndary change
48859	2020.023M.R. Rhabdoviridae_7ngen_16nsp.zip	NEW Zhezhang alphacr ustrhavirus	130390	2020.030M.R. Negamaviricota_correction s.zip	Zhezhang alphacrusthavirus RENAME Zhejiang alphacr ustrhavirus
54855	2020.027M.R. Nairoviridae_4ngen_30nsp.zip	NEW Saphire orthonairovirus	130391	2020.030M.R. Negamaviricota_correction s.zip	Saphire orthonairovirus RENAME Sapphire orthonairovirus
29947	2020.012M.R. Orthobunyavirus_16new_sp_abolish _1sp.zip	NEW Buffalo Creek orth obunyavirus	130392	2020.030M.R. Negamaviricota_correction s.zip	Buffalo Creek orthobunyavirus RENAME Buffalo Creek orth obunyavirus

56844	2020.029M.R. Phenuiviridae_1gen16sp.zip	Guadaloupe phasivirus	130393	2020.030M.R. Negamaviricota_correction s.zip	Guadaloupe phasivirus RENAME • Guadaloupe phasivirus
-------	--	--------------------------	--------	--	--

- N=2 rows are new families where the sheet has the parent order filled in. That's ok, but hard to detect automatically.
 - sort in (92052,113438)
- sort=122400, action=abolish, proposal=2020.169B.R.Tunavirus.zip, *ERROR: no species named*; **Solution:** remove - there's no mention of it in the proposal
- sort=11022, action=move species, proposal=2020.005B.R.Ackermannviridae.zip, *ERROR: species name is "Erwinia virus Ea2810", should be "Erwinia virus Ea2809"*
- sort=1216, action=move subfamily, 2020.001M_014M_015M_016M.R.Rhabdoviridae.zip, *ERROR: no previous subfamily named, appears it should be "create subfamily"*. Changing and proceeding.
 - commit transaction
- **Backup DB** to D:\MSSQL\Backup\ICTVonline36\ICTVonline36_20210430_2035_pre_alter_src_lineage_col.bak

2.d More QC

- **"2.d.QC.check_molecule_types"**
 - N=2370, all fixed with script's guessing logic (remove the space)

report	usage_ct	molecule	molecule_id	best_guess	problem
map molecule names to taxonomy_molecule	336	ssRNA (-)	--MISSING--	ssRNA(-)	ERROR: misssing molecule
map molecule names to taxonomy_molecule	118	ssDNA (-)	--MISSING--	ssDNA(-)	ERROR: misssing molecule
map molecule names to taxonomy_molecule	358	ssDNA (+/-)	--MISSING--	ssDNA(+/-)	ERROR: misssing molecule
map molecule names to taxonomy_molecule	1414	ssRNA (+)	--MISSING--	ssRNA(+)	ERROR: misssing molecule
map molecule names to taxonomy_molecule	1061	dsDNA	1	dsDNA	
map molecule names to taxonomy_molecule	144	ssDNA (+)	--MISSING--	ssDNA(+)	ERROR: misssing molecule
map molecule names to taxonomy_molecule	12	dsDNA-RT	7	dsDNA-RT	
map molecule names to taxonomy_molecule	16	dsRNA	3	dsRNA	

- **"2.h.QC_check_for_re-creates.sql"**
 - sort=88750, proposal=2020.094B.R.Leuconostoc_siphoviruses.zip, change=Create new, _dest_lineage=Duplodnaviria;Heunggongvirae;Uroviricota;Caudoviricetes;Caudovirales;Siphoviridae;McCleskeyvirinae;Unaquatrovirus;Leuconostoc virus 1A4
 - already exists: https://talk.ictvonline.org/taxonomy/p/taxonomy-history?taxnode_id=20175520
 - isWrong=Already exists
- **"2.Y summarize in-out changes.sql"**: no problems
 - out_change

issue	tree_id	out_change	ct
prev MSL out_change summary	201850000	merge	1
prev MSL out_change summary	201850000	abolish	18
prev MSL out_change summary	201850000	rename	51
prev MSL out_change summary	201850000	move	225
prev MSL out_change summary	201850000	promote	1

- in change

issue	tree_id	in_change	ct
current MSL in_chnage summary	201900000	NULL	6877
current MSL in_chnage summary	201900000	new	1593
current MSL in_chnage summary	201900000	split	11

3. Create new MSL in taxonomy_node

- **"2.Z.insert_root_in_taxonomy_node.sql"**
 - edit in details.
- Backup "ICTVonline36_20210430_2110_pre_3.a_create_MSL.bak"
- **"3.a.create_next_MSL_by_copying_previous_MSL.sql"**
- **"3.c.load_next_msl-set-dest_taxnode_id-dest_parent_id.sql"**
 - begin transaction
 - Error report

FIX	ERROR_report	sort	_act ion	_dest taxo n_rank	rank	_dest_lineage	_dest parent _name	_dest_taxo n_name	dest_ taxno de_id	dest_ _par ent_id	des t_ic tv_id
Metaxyviridae: created MSL36.47869 Cofodevirus: never created, line missing; FIX: add MSL36.4769.5 to create Cofodevirus	actions that should have dest_parent_id set, but do NOT	47870	move	species	species	Monodnaviria;Shotokuvirae;Cressdnaviricota; Arfiviricetes;Mulpavirales;Metaxyviridae;Cofodevirus;Coconut foliar decay virus	Cofodevirus	Coconut foliar decay virus	2020 03907	NULL	NULL
add 'promote' to list of actions where one searches parent in load_next_msl	actions that should have dest_parent_id set, but do NOT	78209	promote	subfamily	subfamily	Duplodnaviria;Heunggongvirae;Uroviricota; Caudoviricetes;Caudovirales;Siphoviridae; Hendrixvirinae	Siphoviridae	Hendrixvirinae	2020 05557	NULL	NULL

- add missing genus
- re-run set dest parent_id
- commit transaction
- "4.a.apply_create_actions_RANK_high_to_low.sql"
 - begin transaction
 - fix code to have "move" code handle "promote" and "demote" by doing at time of _dest_taxon_rank
 - error: sort=89237, RENAME class: Riboviria;Orthonavirae;Lenarviricota;Allasoviricetes >>> Riboviria;Orthonavirae;Lenarviricota;Leviviricetes
 - looks like on 20200421 I updated SP [taxonomy_node_compute_indexes] to have 15 OUTPUT arguments between @species_id and @inher_molecule_id, but that the trigger on taxonomy_node didn't get updated for that.
 - <https://github.com/rusalkaguy/ICTVonlineDbLoad/issues/2>
 - fixed: 0.b4.alter_taxonomy_node.TR_taxonomy_node_update_indexes.sql
 - Backup: ICTVonline36_20210503_0042_post_issue2_fix_pre_4.a.apply.bak
 - Error in step MERGE @ genus

```
-----
-- MERGE @ genus
-----
```

Msg 547, Level 16, State 0, Line 308

The DELETE statement conflicted with the SAME TABLE REFERENCE constraint "FK_taxonomy_node_taxonomy_node_genus_id". The conflict occurred in database "ICTVonline36", table "dbo.taxonomy_node", column 'genus_id'.

The statement has been terminated.

TR_taxonomy_node_update_indexes: ROWS: DEL=1, INS=1

Msg 50000, Level 1, State 1

- sort=8342, merge genus Viunavirus (contains 6 subtaxa) into genus Kuttervirus
-
-
- CODE ERROR: does not move child taxa to merge target before deleting mergee.
- <https://github.com/rusalkaguy/ICTVonlineDbLoad/issues/3>
- QC

report	action	prevMSL	load_next_msl	nextMSL	(No column name)
counts by [ACTION]	abolish	NULL	32	NULL	ERROR; but implemented @ later step
counts by [ACTION]	merge	13	13	NULL	OK
counts by [ACTION]	move	204	204	NULL	OK
counts by [ACTION]	new	NULL	3436	3435	ERROR
counts by [ACTION]	promote	1	1	NULL	OK
counts by [ACTION]	rename	191	191	NULL	OK
counts by [ACTION]	split	NULL	NULL	NULL	OK
counts by [ACTION]	type	NULL	NULL	NULL	OK

- sort=66777 2020.042B 2020.042B.R.Deltatectivirus_Epsilontectivirus.zip; this is "new genus" but with no taxon specified.
 - FIX: set isWrong="blank entry": update load_next_msl set isWrong = 'blank entry in MSL' where sort=66777
 - commit transaction
 - backup: ICTVonline36_20210504_0027_post_4.a.apply.bak
- "4.b.apply_delete_actions_RANK_low_to_high.sql"
 - begin transaction
 - correct code for isWrong being a varchar, not an int.
 - commit transaction
- "4c.qc_actions.sql"
 - fix code for prev_msl QC: 34 (@msl-1)
 - QC by action: OK
 - QC by action,rank: error: 1 promote counted in both subfamily and genus (it's a promote, duh!)
 - Go back and fix "move/rename/promote" code in "4.a.apply_create_actions_RANK_high_to_low.sql" to correctly set desintation rank.
 - Add "demote" as a valid action "0.b5.taxonomy_change_out_INSERT_demote.sql"
 - change sort=59299 ; genus=>subgenus to _action='demote'
 - change sort=113438 -- (change=new family); set all srcRANK fields to NULL.
 - change sort= 92052 -- (change=new family); set all srcRANK fields to NULL.

- that way, this query returns only promotes and demotes

```
select * from load_next_msl where isnull(_src_taxon_rank,_dest_taxon_rank) <> _dest_taxon_rank
```

- lots of code fixes to correctly report on actions that change rank levels.
- network problems. Backup db on dev.ictvonline.org and download to genome-ws-02
 - \\genome-ws-02\\E\$\\ICTVonlineDbLoad\\backups\\ICTVonline36_20210504_1135_pre_5summarize..bak
- **"5.summarize_status_by_rank_and_action.sql"**
 - blank out dest" on an isWrong line (sort=66777)
 - improve code to only check load_new_msl.dest_ictv_id if _action in ("new","split")
 - backup: on \\genome-ws-02\\E\$\\ICTVonlineDbLoad\\backups\\ICTVonline36_20210504_1215_post_5summarize..bak"
- problem proposal/filenames (from 2019) - to be dealt with later.
 - 201900160 2019.022M; 2019.026M.zip 2019.022M.2sp_Pacuvirus.xlsx
 - 201907416 2019.014M; 2019.025M.zip 2019.014M.Avulavirus_1newsp.xlsx
 - <https://github.com/rusalkaguy/ICTVonlineDbLoad/issues/8>
- **"7d.QC check for duplicate taxon names.sql"** no issues 😊
- **"7e.QC_check_for_genera_with_type_species_issues.sql"** no issues 😊
- **"7f.QC check for ictv_id wo a new change-proposal.sql"** - no (new) issues
 - last ICTV_id created w/o a documented "in_change=new/split" was MSL29.20140192 Caudovirales;Podoviridae;F116likevirus;Pseudomonas phage F116
- **"7g.QC_molecule_id.sql"**
 - (sort= 104496) - wrong molecule? Typo?

Duplodnaviria;Heunggongvirae;Uroviricota;Caudoviricetes;Caudovirales;Myoviridae;Rahariannevirus

Is tagged with dsRNA, while it's parent taxa have dsDNA, and it's member species

Duplodnaviria;Heunggongvirae;Uroviricota;Caudoviricetes;Caudovirales;Myoviridae;Rahariannevirus;Ralstonia virus
Raharianne

Has dsDNA.

- Many genera in the Phylum Riboviria;Orthornavirae;Kitrinoviricota don't have a molecule type.
- (sort=52861) ssDNA(-) while all around is ssRNA(-): genus Culicidavirus
- Many genera in the class Riboviria;Orthornavirae;Pisuviricota;Pisoniviricetes don't have a molecule type
- Riboviria;Orthornavirae;Pisuviricota;Stelpaviricetes
- Naldaviricetes;Lefavirales;Baculoviridae;Alphabaculovirus;Spodoptera exigua multiple nucleopolyhedrovirus A/B are dsDNA while other species in genus are dsDNA-RT
- Created report and sent to Elliot: **"7g.REPORT.QC_molecule_id.xlsx"**
- **Need to return and fix**
- **"8a.rebuild_delta_nodes_and_merge-split_table.sql"**
 - no - we need to run this again after handling isType issue
- **"7h. QC merge split table.sql"** - no errors.
- **"8c. pull isolates-accession-molecule.sql"**
 - Should this be before **"7g.QC_molecule_id.sql"**?
- **"8d. QC isolates-accession-molecule.sql"**
 - document changes in **"8d. QC isolates-accession-molecule.report.xlsx"** & emailed to Elliot
- **"9a. recompile views, etc.sql"**: done
- **"9x.QC_deltas.sql"**
 - ERROR DETAIL: MSL36 tax APPEARS with out a delta
record NULL NULL 36 NULL NULL 202005054 20143656 family Sphaerolipoviridae Varidnaviria;Helvetiavirae;
Dividoviricota;Laserviricetes;Halopanivirales;Sphaerolipoviridae
 - sort=92052 was 'new family' but had src=Varidnaviria;Helvetiavirae;Dividoviricota;Laserviricetes;Halopanivirales;
Sphaerolipoviridae - so it became a move and messed things up
should have created 'Halopanivirales;[202011533]Matshushitaviridae', instead renamed '[201905054]Sphaerolipoviridae '
but also created [202005054]Sphaerolipoviridae
 - removed in_target='Sphaerolipoviridae' in **"9x.QC_deltas.xFix2_MSL36_Sphaerolipoviridae.sql"**
 - ERROR DETAIL: MSL36 tax APPEARS with out a delta
record NULL NULL 36 NULL NULL 202008706 201908706 order Halopanivirales Varidnaviria;Helvetiavirae;
Dividoviricota;Laserviricetes;Halopanivirales
 - '[202008706]' - lacks IN delta
 - '[201908706]Halopanivirales' => new 'Varidnaviria;Helvetiavirae;Dividoviricota;Laserviricetes;Halopanivirales;Simuloviridae'
 - sort=113438 was a 'new famil', but has src=Varidnaviria;Helvetiavirae;Dividoviricota;Laserviricetes;Halopanivirales - so it became a messed-up move-new-thing
 - fixed in **"9x.QC_deltas.xFix3_MSL36_Halopanivirales.sql"**
- **BACKUP: "ICTVonline36_20210505_0037_post_9x.QC_deltas.xFix3.bak"**
 - 9x.QC_deltas.xFix4_MSL36_Rahariannevirus_mol.sql
- **"9za.EXPORT_MSL_extended_from_taxonomy_node.sql"**
 - exec MSL_export_official --11m31s on genome-ws-02
- FEATURE-to-do: taxonomy_node_delta.is_now_type - obsolete this
 - <https://github.com/rusalkaguy/ICTVonlineDbLoad/issues/5>
 - **"E:ICTVonlineDbLoad\\9x.QC_deltas.FixMSL36.5_rebuild_delta_nodes_SP.sql"**
- **BACKUP: "ICTVonline36_20210505_1348_post_9x.FixMSL36.5_is_type.bak"**
- FEATURE-to-do: taxonomy_node_delta.is_demote - add this
 - <https://github.com/rusalkaguy/ICTVonlineDbLoad/issues/4>
 - main script: **"9x.QC_deltas.FixMSL36.6_demote_support.sql"**
 - add column taxonomy_node_delta.is_demoted

- update SP rebuild_delta_nodes to compute it "9x.QC_deltas.FixMSL36.6_rebuild_delta_nodes_SP.sql"
- tags_csv schema script: "9x.QC_deltas.FixMSL36.6_taxonomy_node.ALTER.tags_csv.sql"
 - tag_csv is a computed (but indexed) column, so
 - drop index
 - drop column (can't alter computed columns in MSSQL)
 - create new column
 - re-create index
- BACKUP: "ICTVonline36_20210505_1348_post_9x.FixMSL36.6_is_demoted.bak"
- Final counts cross-check with [Elliot Lefkowitz](#)
 - select msl_release_num, taxnode_id, taxa_desc_cts from taxonomy_node where level_id =100 and msl_release_num=36
 - 6 realms, 10 kingdoms, 17 phyla, 2 subphyla, 39 classes, 59 orders, 8 suborders, 189 families, 136 subfamilies, 2224 genera, 70 subgenera, 9110 species
 - Missing 1 subfamily
 - sort=1216; new subfamily coded as moved; lineage=Riboviria;Orthornavirae;Negarnaviricota;Haploviricotina;Monjiviricetes; Mononegavirales;Rhabdoviridae;Gammarhabdovirinae
 - [Elliot Lefkowitz](#) was tracing down a discrepancy between the number of new subfamilies, and it came down to a miscoding of the creation of 'Gammarhabdovirinae':

1. 2020.016M.R.Rhabdoviridae_3subfam.zip
 - a. Problem: The creation of the subfamily 'Gammarhabdovirinae' was coded as a 'move' in 2020.001 M_014M_015M_016M.R.Rhabdoviridae; should be a 'Create new'
 - b. Solution: Recode as a 'Create new'

It looks like your code fixed this. Did you report this previously?

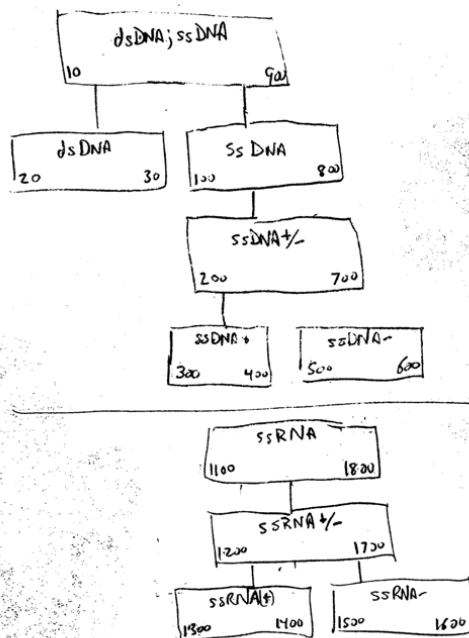
- [Curtis Hendrickson](#) Doesn't look like we discussed that one.

Looks like the code that translates from the "load_next_msl.change" column, which is populated from the spread sheet, to the "load_next_msl.action" column, which is used for data integration, caught and corrected that on its own.

I've added several heuristics over the years, so it treats "change" as more guidance than fact, sometimes.

Sorry I didn't notice and report it.

- FEATURE-to-do: taxonomy_node_delta.is_now_type - just un_is_ref thes pecies, restore the machinery
 - <https://github.com/rusalkaguy/ICTVonlineDbLoad/issues/5>
 - "9x.QC_deltas.FixMSL36.7_proposal_for_is_type.sql"
 - N:1421 where proposal like '%Abolish_type_species%'
- Molecule cleanups
 - push up species higher where consistent (no further than order)
 - in bunyavirales, never above genus
 - "exec sp_simplify_molecule_id_settings"
 - already limits to ORDER and below
 - added code to enforce genus-only in Bunyavirales (20210509 re-added; had to rollback to last backup).
 - look at Bunyavirales - lots of things set on family....
 - select taxnode_id, [rank], name, molecule, inher_molecule from taxonomy_node_names where [order] = 'Bunyavirales' and msl_release_num=36 order by left_idx
 - heuristic that ssDNA(*) is not a contradiction of ssDNA on a higher rank, just more precision. And ssDNA(-) is a subtype of ssDNA(+/-), etc.
 - Issue #6: <https://github.com/rusalkaguy/ICTVonlineDbLoad/issues/6>
 - Hand-edit taxonomy_molecule in GUI to add left_idx & right_idx columns
 - Use Nested Sets to store queryable hierarchy: <https://docs.microsoft.com/en-us/archive/blogs/mvpawardprogram/hierarchies-convert-adjacency-list-to-nested-sets#nested-sets>



- Hand-edit taxonomy_molecule to set left_idx, right_idx values

GENOME-WS-02.IC...taxonomy_molecule				dbo.sp_simplify_mol...(UAB\curtish (59))			9x.QC_delta	
	id	abbrev	name	balt_gro...	balt_rom...	descripti...	left_idx	right_idx
	0	Unassigned	Unassign...	NULL	NULL	NULL	NULL	NULL
	1	dsDNA	Double-s...	I	I	NULL	20	30
	2	ssDNA	Single-str...	II	II	NULL	100	800
	3	dsRNA	Double-s...	III	III	NULL	NULL	NULL
	4	ssRNA(+)	Single-str...	IV	IV	NULL	1300	1400
	5	ssRNA(-)	Single-str...	V	V	NULL	1500	1600
▶	6	ssRNA-RT	Single-str...	VI	VI	NULL	NULL	NULL
	7	dsDNA-RT	Double-s...	VII	VII	NULL	NULL	NULL
	8	Viroid	Circular S...	NULL	NULL	NULL	NULL	NULL
	9	ssDNA(-)	Single-str...	NULL	NULL	NULL	500	600
	10	ssDNA(+)	Single-str...	NULL	NULL	NULL	300	400
	11	ssDNA(+/-)	Single-str...	NULL	NULL	NULL	200	700
	12	ssRNA(+/-)	Single-str...	NULL	NULL	NULL	1200	1700
	13	dsDNA; ssDNA	DNA - so...	NULL	NULL	NULL	10	900
	14	ssRNA	Single-str...	NULL	NULL	Introduc...	1100	1800
*	NULL	NULL	NULL	NULL	NULL	NULL	NULL	NULL

- send update qc report to elliot for typos. N=45 "overrides": "7g.REPORT.QC_molecule_id.xlsx"
- elliot sent back updates: "7g.REPORT.QC_molecule_id-EJL.xlsx"
- "E:\ICTVonlineDbLoad\9x.QC_deltas.FixMSL36.8_fix_molecule_ids.sql"
 - problem 1: 202004409 (family: Polyomaviridae) is Incorrectly marked ssDNA, when it should be dsDNA; species can be set to NULL and inherit. (all new species were added with dsDNA and conflicted)
 - BACKUP:** "ICTVonline36_20210509_2117 post_Polyomaviridae_molecule_fixMSL36.8.bak"
 - problem 2: Pleolipoviridae;...;Alphapleolipovirus HHPV1 (dsDNA!=ssDNA(+/-)!=ssDNA
 - In Pleolipoviridae, the species in red was added with mol=dsDNA. The family is ssDNA, and that genus is mostly ssDNA(+/-).

So, one species doesn't match the rest of the genus.

You commented

"Viruses in the Pleolipoviridae family can be either dsDNA or ssDNA; HHPV1 is dsDNA; HRPV1 and HRPV2 are ssDNA; Most dsDNA; HRPV-3 and HGPV-1 are both: 'dsDNA; ssDNA'"

However, but HHPV1 and HRPV1 are in not just the same family, but the same genus. And currently HHPV2 matches HRPV1/2/6 (ssDNA(+/-) – see attached image

- -- query local tree
- DECLARE @name varchar(50); SET @name='Pleolipoviridae'
- select n.taxnode_id, n.ictv_id, n.[rank], n.molecule, n.lineage, n.*
- from taxonomy_node_names n
- join taxonomy_node t on n.left_idx between t.left_idx and t.right_idx and n.tree_id = t.tree_id
- where t.msl_release_num=36
- and t.name =@name
- order by n.left_idx

I'm guessing I should make

1. Alphapleolipovirus genus, and all its species, either ssDNA or ssDNA(+/-)
 - a. Or make HHPV1/2 dsDNA and HRPV1/2/6 ssDNA(+/-).
2. Leave Betapleolipovirus alone
3. Remove any molecule setting from the family Pleolipoviridae

taxnode_id	ictv_id	rank	molecule	lineage	taxnode_id	parent_id	tree_id
1	202004340	20154620	family	ssDNA	202004340	202007184	202000000
2	202004341	20154621	genus	ssDNA	202004341	202004340	202000000
3	202004342	20154625	species	ssDNA	202004342	202004341	202000000
4	202004343	20154626	species	ssDNA(+/-)	202004343	202004341	202000000
5	202004344	20154622	species	ssDNA(+/-)	202004344	202004341	202000000
6	202004345	20154623	species	ssDNA(+/-)	202004345	202004341	202000000
7	202004346	20154624	species	ssDNA(+/-)	202004346	202004341	202000000
8	202004347	20154627	genus	ssDNA	202004347	202004340	202000000
9	202004348	20154629	species	ssDNA	202004348	202004347	202000000
10	202008635	201908635	species	ssDNA	202008635	202004347	202000000
11	202008636	201908636	species	ssDNA	202008636	202004347	202000000
12	202004349	20154628	species	ssDNA	202004349	202004347	202000000
13	202008637	201908637	species	ssDNA	202008637	202004347	202000000
14	202008638	201908638	species	ssDNA	202008638	202004347	202000000
15	202008639	201908639	species	ssDNA	202008639	202004347	202000000
16	202008640	201908640	species	ssDNA	202008640	202004347	202000000
17	202008641	201908641	species	ssDNA	202008641	202004347	202000000
18	202004350	20154630	genus	ssDNA	202004350	202004340	202000000
19	202004351	20154631	species	ssDNA	202004351	202004350	202000000

- Here is what should be done:

- Remove mol type from the family Pleolipoviridae
- Betapleolipovirus are dsDNA except HRPV-* and HGPV-* are 'dsDNA; ssDNA'
- Gammapleolipovirus is dsDNA
- Alphapleolipovirus are mixed: All ssDNA except HHPV-1 is dsDNA

report	taxnode_id	rank	molecule	inher_molecule	lineage
1	report on molecule 202007161	realm	ssDNA	ssDNA	Monodnaviria
2	report on molecule 202007181	kingdom	ssDNA	ssDNA	Monodnaviria, Trapvirae
3	report on molecule 202007182	phylum	ssDNA	ssDNA	Monodnaviria, Trapvirae, Salevnicota
4	report on molecule 202007183	class	ssDNA	ssDNA	Monodnaviria, Trapvirae, Salevnicota, Huolimaviricetes
5	report on molecule 202007184	order	ssDNA	ssDNA	Monodnaviria, Trapvirae, Salevnicota, Huolimaviricetes, Halourvirales
6	report on molecule 202004340	family	ssDNA	ssDNA	Monodnaviria, Trapvirae, Salevnicota, Huolimaviricetes, Halourvirales, Pleolipoviridae
7	report on molecule 202004341	genus	ssDNA	ssDNA	Monodnaviria, Trapvirae, Salevnicota, Huolimaviricetes, Halourvirales, Pleolipoviridae, Alphapleolipovirus
8	report on molecule 202004342	species	dsDNA	dsDNA	Monodnaviria, Trapvirae, Salevnicota, Huolimaviricetes, Halourvirales, Pleolipoviridae, Alphapleolipovirus, HHPV1
9	report on molecule 202004343	species	dsDNA	dsDNA	Monodnaviria, Trapvirae, Salevnicota, Huolimaviricetes, Halourvirales, Pleolipoviridae, Alphapleolipovirus, HHPV2
10	report on molecule 202004344	species	ssDNA	ssDNA	Monodnaviria, Trapvirae, Salevnicota, Huolimaviricetes, Halourvirales, Pleolipoviridae, Alphapleolipovirus, HHPV3
11	report on molecule 202004345	species	ssDNA	ssDNA	Monodnaviria, Trapvirae, Salevnicota, Huolimaviricetes, Halourvirales, Pleolipoviridae, Alphapleolipovirus, HHPV4
12	report on molecule 202004346	species	ssDNA	ssDNA	Monodnaviria, Trapvirae, Salevnicota, Huolimaviricetes, Halourvirales, Pleolipoviridae, Alphapleolipovirus, HHPV5
13	report on molecule 202004347	genus	dsDNA, ssDNA	dsDNA, ssDNA	Monodnaviria, Trapvirae, Salevnicota, Huolimaviricetes, Halourvirales, Pleolipoviridae, Betapleolipovirus
14	report on molecule 202004348	species	dsDNA, ssDNA	dsDNA, ssDNA	Monodnaviria, Trapvirae, Salevnicota, Huolimaviricetes, Halourvirales, Pleolipoviridae, Betapleolipovirus, HGPV1
15	report on molecule 202008635	species	dsDNA	dsDNA	Monodnaviria, Trapvirae, Salevnicota, Huolimaviricetes, Halourvirales, Pleolipoviridae, Betapleolipovirus, HHPV3
16	report on molecule 202008636	species	dsDNA	dsDNA	Monodnaviria, Trapvirae, Salevnicota, Huolimaviricetes, Halourvirales, Pleolipoviridae, Betapleolipovirus, HHPV4
17	report on molecule 202004349	species	dsDNA, ssDNA	dsDNA, ssDNA	Monodnaviria, Trapvirae, Salevnicota, Huolimaviricetes, Halourvirales, Pleolipoviridae, Betapleolipovirus, HHPV9
18	report on molecule 202008637	species	dsDNA, ssDNA	dsDNA, ssDNA	Monodnaviria, Trapvirae, Salevnicota, Huolimaviricetes, Halourvirales, Pleolipoviridae, Betapleolipovirus, HHPV10
19	report on molecule 202008638	species	dsDNA, ssDNA	dsDNA, ssDNA	Monodnaviria, Trapvirae, Salevnicota, Huolimaviricetes, Halourvirales, Pleolipoviridae, Betapleolipovirus, HHPV11
20	report on molecule 202008639	species	dsDNA, ssDNA	dsDNA, ssDNA	Monodnaviria, Trapvirae, Salevnicota, Huolimaviricetes, Halourvirales, Pleolipoviridae, Betapleolipovirus, HHPV12
21	report on molecule 202008640	species	dsDNA	dsDNA	Monodnaviria, Trapvirae, Salevnicota, Huolimaviricetes, Halourvirales, Pleolipoviridae, Betapleolipovirus, HRPV1
22	report on molecule 202008641	species	dsDNA	dsDNA	Monodnaviria, Trapvirae, Salevnicota, Huolimaviricetes, Halourvirales, Pleolipoviridae, Betapleolipovirus, HRPV2
23	report on molecule 202004350	genus	dsDNA	dsDNA	Monodnaviria, Trapvirae, Salevnicota, Huolimaviricetes, Halourvirales, Pleolipoviridae, Gammapleolipovirus
24	report on molecule 202004351	species	dsDNA	dsDNA	Monodnaviria, Trapvirae, Salevnicota, Huolimaviricetes, Halourvirales, Pleolipoviridae, Gammapleolipovirus, Hs2

- Nope - clarification

- The last screen shot was not right. The species designations should be exactly as indicated below. I would remove all values for mol type at all higher ranks. Just assign to species as follows:

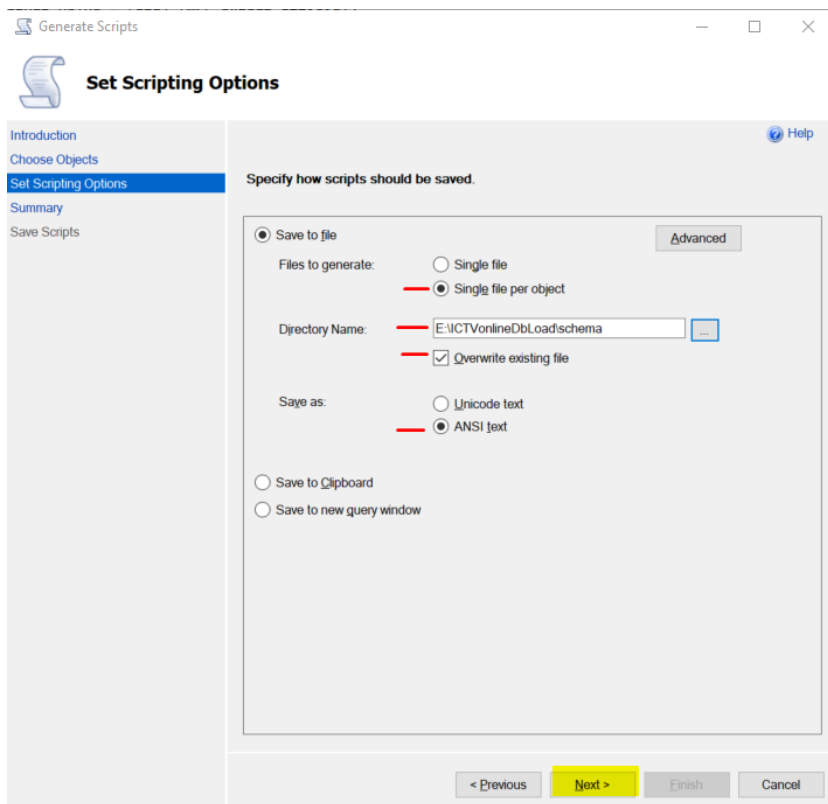
1. Alphapleolipovirus are mixed: All ssDNA except HHPV-1 is dsDNA
2. Betapleolipovirus are dsDNA except HRPV-3 and HGPV-1 are 'dsDNA; ssDNA'
3. Gammapleolipovirus is dsDNA
4. "9x.QC_deltas.FixMSL36.8_fix_molecule_ids.sql"

	report	taxnode_id	rank	molecule	inher_molecule	name	lineage
1	report on molecule	202007161	realm	ssDNA	ssDNA	Monodnaviria	Monodnaviria
2	report on molecule	202007181	kingdom		ssDNA	Trapavirae	Monodnaviria;Trapavirae
3	report on molecule	202007182	phylum		ssDNA	Saleviricota	Monodnaviria;Trapavirae;Saleviricota
4	report on molecule	202007183	class		ssDNA	Huolimaviricetes	Monodnaviria;Trapavirae;Saleviricota;Huolimaviri...
5	report on molecule	202007184	order		ssDNA	Haloruvirales	Monodnaviria;Trapavirae;Saleviricota;Huolimaviri...
6	report on molecule	202004340	family		ssDNA	Pleolipoviridae	Monodnaviria;Trapavirae;Saleviricota;Huolimaviri...
7	report on molecule	202004341	genus		ssDNA	Alphapleolipovirus	Monodnaviria;Trapavirae;Saleviricota;Huolimaviri...
8	report on molecule	202004342	species	dsDNA	dsDNA	Alphapleolipovirus HHPV1	Monodnaviria;Trapavirae;Saleviricota;Huolimaviri...
9	report on molecule	202004343	species	ssDNA	ssDNA	Alphapleolipovirus HHPV2	Monodnaviria;Trapavirae;Saleviricota;Huolimaviri...
10	report on molecule	202004344	species	ssDNA	ssDNA	Alphapleolipovirus HRPV1	Monodnaviria;Trapavirae;Saleviricota;Huolimaviri...
11	report on molecule	202004345	species	ssDNA	ssDNA	Alphapleolipovirus HRPV2	Monodnaviria;Trapavirae;Saleviricota;Huolimaviri...
12	report on molecule	202004346	species	ssDNA	ssDNA	Alphapleolipovirus HRPV6	Monodnaviria;Trapavirae;Saleviricota;Huolimaviri...
13	report on molecule	202004347	genus	dsDNA; ssDNA	dsDNA; ssDNA	Betapleolipovirus	Monodnaviria;Trapavirae;Saleviricota;Huolimaviri...
14	report on molecule	202004348	species	dsDNA; ssDNA	dsDNA; ssDNA	Betapleolipovirus HGPV1	Monodnaviria;Trapavirae;Saleviricota;Huolimaviri...
15	report on molecule	202008635	species	dsDNA	dsDNA	Betapleolipovirus HHPV3	Monodnaviria;Trapavirae;Saleviricota;Huolimaviri...
16	report on molecule	202008636	species	dsDNA	dsDNA	Betapleolipovirus HHPV4	Monodnaviria;Trapavirae;Saleviricota;Huolimaviri...
17	report on molecule	202004349	species	dsDNA; ssDNA	dsDNA; ssDNA	Betapleolipovirus HRPV3	Monodnaviria;Trapavirae;Saleviricota;Huolimaviri...
18	report on molecule	202008637	species	dsDNA	dsDNA	Betapleolipovirus HRPV9	Monodnaviria;Trapavirae;Saleviricota;Huolimaviri...
19	report on molecule	202008638	species	dsDNA	dsDNA	Betapleolipovirus HRPV10	Monodnaviria;Trapavirae;Saleviricota;Huolimaviri...
20	report on molecule	202008639	species	dsDNA	dsDNA	Betapleolipovirus HRPV11	Monodnaviria;Trapavirae;Saleviricota;Huolimaviri...
21	report on molecule	202008640	species	dsDNA	dsDNA	Betapleolipovirus HRPV12	Monodnaviria;Trapavirae;Saleviricota;Huolimaviri...
22	report on molecule	202008641	species	dsDNA	dsDNA	Betapleolipovirus SNJ2	Monodnaviria;Trapavirae;Saleviricota;Huolimaviri...
23	report on molecule	202004350	genus	dsDNA	dsDNA	Gammappleolipovirus	Monodnaviria;Trapavirae;Saleviricota;Huolimaviri...
24	report on molecule	202004351	species	dsDNA	dsDNA	Gammappleolipovirus His2	Monodnaviria;Trapavirae;Saleviricota;Huolimaviri...

- Coguvirid genus should all be ssRNA(+/-). Otherwise, I would make no other changes to Mol type for taxa within the order Bunyavirales.
 - "9x.QC_deltas.FixMSL36.8_fix_molecule_ids.sql"
- Duplodnaviria;Heunggongvirae;Uroviricota;Caudoviricetes;Caudovirales;Podoviridae;Firingavirus;Ralstonia virus RSK1 should be dsDNA
 - "9x.QC_deltas.FixMSL36.8_fix_molecule_ids.sql"
 - **BACKUP:** "ICTVonline36_20210510_0755 post_fixMSL36.8.bak"
- Elliot updates on report names:
 - in solving missing creation of family Metaxyviridae

action	original MSL	my original hack	Elliot MSL solution
create family Metaxyviridae	missing	47869.5	47868
create genus Cofodovirus	47869	47869	47869

- first update db 'sort' values to match Elliot (load_next_msl.sort & taxonomy_node.in_notes)
- update load_next_msl from spreadsheet via generated SQL updates
- then update taxonomy_node: out_filename & in_filename
- re-apply "abolish type species" proposal
- QC checks
 - 89241 2020.095B.R.Leviviricetes-orig.zip;2020.001G.R.Abolish_type_species.pdf 2020.095B.R.Leviviricetes.zip
 - Riboviria;Orthornavirae;Lenarviricota;Allasoviricetes;Levivirales;Leviviridae;Levivirus;Escherichia virus MS2
- save schema, fix and git commit
 - save schema
 - right-click no ICTVonline36 > Tasks > Generate Scripts ...



-
- Remove db name and dates
 - cd schema
 - ./cleanup_sql.sh
 - git add -u .
 - git add X Y Z....
 - git commit
- update the release description to: "EC 52, Online meeting, October 2020; Email ratification March 2021 (MSL #36)"
 - "9x.QC_deltas.FixMSL36.A3_update_MSL_description.sql"
- update MSL exportSP
 - MSL export included a column named "Last Change Rank" that was confusing, but also showed there was a problem with molecule annotations.
 - Removed "Last Change Rank" and "IsTypeSpecies" from MSL_export_official SP
 - "9x.QC_deltas.FixMSL36.A2_alter_MSL_export_official.sql"
 - "9x.QC_deltas.FixMSL36.A2_create_MSL_export_official_preMSL36.sql"
 - Something got messed up in MSL35 molecules - so we had many taxa missing molecule_id settings, that the MSL exporter was able to find the molecules in older MSLs.
 - implemented an update to pull most recent history molecule ID forward into MSL36. Pulled species, then ran the simplification SP to push annotations up the ranks.
 - "9x.QC_deltas.FixMSL36.A1_pull_molecules_from_prev.sql"
- Final backup for production
 - Elliot Ran the "9x.QC_deltas.FixMSL36.A*.sql" on production, plus running "exec rebuild_delta_nodes". That should be equivalent to this backup.
 - **BACKUP:** "ICTVonline36_20210510_1723_post_fixMSL36.A_and_delta_rebuild.bak", which I copied to box:/ictvonline/taxonomy/ICTV_update/2020_updates/
-