Dfam 3.0 Schema

repeatmasker_stage
- 🔑 id BIGINT(20)
- ◇ name VARCHAR(25)
- ◇ description VARCHAR(128)
- Indexes

family_has_search_stage
- family_id BIGINT(20)
- repeatmasker_stage_id BIGINT(20)
- Indexes

family_database_alias
- family_id BIGINT(20)
- 🔑 db_id VARCHAR(80)
- 🔑 db_link VARCHAR(80)
- ◇ deprecated BOOLEAN
- ◇ comment TEXT
- Indexes

family_has_citation
- family_id BIGINT(20)
- citation_pmid INT(10)
- ◇ comment TINYTEXT
- ◇ order_added TINYINT(4)
- Indexes

citation
- 🔑 pmid INT(10)
- ◇ title TINYTEXT
- ◇ authors MEDIUMTEXT
- ◇ journal TINYTEXT
- ◇ pubdate DATE
- Indexes

family_has_buffer_stage
- family_id BIGINT(20)
- repeatmasker_stage_id BIGINT(20)
- 🔑 start_pos MEDIUMINT(8)
- 🔑 end_pos MEDIUMINT(8)
- Indexes

seed_align_data
- family_id BIGINT(20)
- ◇ graph_json LONGBLOB
- Indexes

family
- 🔑 id BIGINT(20)
- ◇ accession VARCHAR(45)
- ◇ version SMALLINT(5)
- ◇ name VARCHAR(45)
- ◁ classification_id BIGINT(20)
- ◇ description TEXT
- ◇ consensus TEXT
- ◇ date_created DATETIME
- ◇ date_modified DATETIME
- ◇ date_deleted DATETIME
- ◇ target_site_cons VARCHAR(30)
- ◇ author TINYTEXT
- ◁ deposited_by_id INT(11)
- ◁ curation_state_id BIGINT(20)
- ◇ disabled TINYINT(1)
- ◇ refineable TINYINT(1)
- ◇ model_consensus TEXT
- ◇ model_mask TEXT
- ◇ hmm_build_method_id BIGINT(20)
- ◇ cons_build_method_id BIGINT(20)
- ◇ length INT(11)
- ◇ hmm_maxl INT(11)
- ◇ hmm_general_threshold DOUBLE
- ◇ seed_ref MEDIUMTEXT
- ◇ title VARCHAR(80)
- ◇ curation_notes TEXT
- Indexes
- Triggers
- BEF INSERT family_before_ins_trig

curation_state
- 🔑 id BIGINT(20)
- ◇ name VARCHAR(45)
- ◇ description VARCHAR(255)
- Indexes

wikipedia
- 🔑 auto_wiki INT(10)
- ◇ title TINYTEXT
- ◇ wikitext LONGTEXT
- Indexes

classification_has_wikipedia
- classification_id BIGINT(20)
- auto_wiki INT(10)
- Indexes

family_overlap
- 🔑 id BIGINT(20)
- ◁ family1_id BIGINT(20)
- ◁ family2_id BIGINT(20)
- Indexes

overlap_segment
- family_overlap_id BIGINT(20)
- 🔑 family1_start MEDIUMINT(8)
- 🔑 family1_end MEDIUMINT(8)
- 🔑 family2_start MEDIUMINT(8)
- 🔑 family2_end MEDIUMINT(8)
- 🔑 strand ENUM('+', '-')
- ◇ evalue VARCHAR(15)
- ◇ identity VARCHAR(6)
- ◇ coverage VARCHAR(6)
- ◇ cigar TEXT
- Indexes

seed_region
- ◁ family_id BIGINT(20)
- ◁ assembly_id BIGINT(20)
- ◇ seq_id VARCHAR(50)
- ◇ seq_start BIGINT(20)
- ◇ seq_end BIGINT(20)
- ◇ a3m_seq TEXT
- ◇ strand ENUM('+','-')
- ◇ model_start MEDIUMINT(8)
- ◇ model_end MEDIUMINT(8)
- Indexes

family_clade
- family_id BIGINT(20)
- dfam_taxdb_tax_id BIGINT(20)
- Indexes

assembly
- 🔑 id BIGINT(20)
- ◇ name VARCHAR(45)
- ◇ description VARCHAR(100)
- ◁ dfam_taxdb_tax_id BIGINT(20)
- ◇ source ENUM(...)
- ◇ release_date DATETIME
- ◇ version VARCHAR(45)
- ◇ schema_name VARCHAR(45)
- ◇ visible INT
- ◇ display_order INT
- ◇ z_size BIGINT(20)
- Indexes

dfam_taxdb
- 🔑 tax_id BIGINT(20)
- ◇ scientific_name VARCHAR(128)
- ◇ sanitized_name VARCHAR(128)
- ◇ common_name VARCHAR(128)
- ◇ unique_name VARCHAR(128)
- ◇ lineage MEDIUMTEXT
- Indexes

family_feature
- 🔑 id BIGINT(20)
- ◁ family_id BIGINT(20)
- ◇ feature_type VARCHAR(45)
- ◇ description TEXT
- ◇ model_start_pos INT
- ◇ model_end_pos INT
- ◇ label VARCHAR(45)
- Indexes

classification **Adjacency-List**
- 🔑 id BIGINT(20)
- ◁ parent_id BIGINT(20)
- ◇ name VARCHAR(80)
- ◇ tooltip VARCHAR(128)
- ◇ description TEXT
- ◇ hyperlink VARCHAR(255)
- ◁ repeatmasker_type_id BIGINT(20)
- ◁ repeatmasker_subtype_id BIGINT(20)
- ◇ sort_order INT(8)
- ◇ repbase_equiv VARCHAR(255)
- ◇ wicker_equiv VARCHAR(255)
- ◇ curcio_derbyshire_equiv VARCHAR(255)
- ◇ piegu_equiv VARCHAR(255)
- ◇ lineage TEXT
- ◇ aliases TEXT
- Indexes

hmm_model_data
- family_id BIGINT(20)
- ◇ hmm_logo LONGBLOB
- ◇ hmm LONGBLOB
- Indexes

family_assembly_data
- family_id BIGINT(20)
- assembly_id BIGINT(20)
- ◇ cons_genome_avg_kimura_div_GA FLOAT
- ◇ cons_genome_avg_kimura_div_TC FLOAT
- ◇ hmm_genome_avg_kimura_div_GA FLOAT
- ◇ hmm_genome_avg_kimura_div_TC FLOAT
- ◇ cons_GA_hit_count BIGINT(20)
- ◇ cons_TC_hit_count BIGINT(20)
- ◇ cons_GA_nrph_hit_count BIGINT(20)
- ◇ cons_TC_nrph_hit_count BIGINT(20)
- ◇ hmm_GA_hit_count BIGINT(20)
- ◇ hmm_TC_hit_count BIGINT(20)
- ◇ hmm_GA_nrph_hit_count BIGINT(20)
- ◇ hmm_TC_nrph_hit_count BIGINT(20)
- ◇ hmm_hit_GA DOUBLE
- ◇ hmm_hit_GA_evalue VARCHAR(15)
- ◇ hmm_hit_TC DOUBLE
- ◇ hmm_hit_TC_evalue VARCHAR(15)
- ◇ hmm_hit_NC DOUBLE
- ◇ hmm_hit_NC_evalue VARCHAR(15)
- ◇ hmm_fdr DOUBLE
- ◇ hmm_method_id INT(10)
- ◇ cons_fdr DOUBLE
- ◇ cons_method_id INT(10)
- ◇ cons_35GC_GA INT
- ◇ cons_37GC_GA INT
- ◇ cons_39GC_GA INT
- ◇ cons_41GC_GA INT
- ◇ cons_43GC_GA INT
- ◇ cons_45GC_GA INT
- ◇ cons_47GC_GA INT
- ◇ cons_49GC_GA INT
- ◇ cons_51GC_GA INT
- ◇ cons_53GC_GA INT
- ◇ cons_matrix_div INT
- ◇ hmm_avg_hit_length INT
- ◇ cons_avg_hit_length INT
- ◇ hmm_thresh_search_evalue VARCHAR(15)
- Indexes

feature_attribute
- family_feature_id BIGINT(20)
- 🔑 attribute VARCHAR(45)
- ◇ value VARCHAR(45)
- Indexes

coding_sequence
- 🔑 id BIGINT(20)
- ◇ product VARCHAR(45)
- ◇ translation TEXT
- ◇ cds_start INT(10)
- ◇ cds_end INT(10)
- ◇ exon_count INT(10)
- ◇ exon_starts LONGBLOB
- ◇ exon_ends LONGBLOB
- ◇ family_id BIGINT(20)
- ◇ external_reference VARCHAR(128)
- ◇ reverse BOOLEAN
- ◇ stop_codons INT
- ◇ frameshifts INT
- ◇ gaps INT
- ◇ percent_identity FLOAT
- ◇ left_unaligned INT
- ◇ right_unaligned INT
- ◁ classification_id BIGINT(20)
- ◇ align_data TEXT
- ◇ description TEXT
- ◇ protein_type VARCHAR(45)
- Indexes

repeatmasker_subtype
- 🔑 id BIGINT(20)
- ◇ name VARCHAR(25)
- ◇ description VARCHAR(128)
- ◇ parent_type_id BIGINT(20)
- Indexes

repeatmasker_type
- 🔑 id BIGINT(20)
- ◇ name VARCHAR(25)
- ◇ description VARCHAR(128)
- Indexes

ncbi_taxdb_names
- 🔑 tax_id BIGINT(20)
- 🔑 name_txt VARCHAR(128)
- 🔑 unique_name VARCHAR(128)
- 🔑 name_class VARCHAR(128)
- ◇ sanitized_name VARCHAR(128)
- Indexes

db_version
- 🔑 dfam_version VARCHAR(45)
- 🔑 dfam_release_date DATE
- Indexes

next_accession
- ◇ next_acc_id BIGINT(20)

dead_family
- 🔑 accession VARCHAR(45)
- ◇ name VARCHAR(45)
- ◇ comment MEDIUMTEXT
- ◇ user INT(11)
- ◇ deleted DATETIME
- Indexes

ncbi_gencode
- 🔑 genetic_code_id INT
- ◇ abbreviation VARCHAR(40)
- ◇ name VARCHAR(80)
- ◇ cde VARCHAR(80)
- ◇ starts VARCHAR(80)
- Indexes

ncbi_taxdb_nodes
- 🔑 tax_id BIGINT(20)
- ◇ parent_id BIGINT(20)
- ◇ rank VARCHAR(128)
- ◇ embl_code VARCHAR(4)
- ◇ division_id BIGINT(20)
- ◇ inherited_div TINYINT(1)
- ◇ genetic_code_id BIGINT(20)
- ◇ inherited_GC TINYINT(1)
- ◇ mitochondrial_genetic_code_id BIGINT(20)
- ◇ inherited_MGC_flag TINYINT(1)
- ◇ GenBank_hidden_flag TINYINT(1)
- ◇ hidden_subtree_root_flag TINYINT(1)
- ◇ comments BLOB
- Indexes