Assembly DB 3.5 Schema

version
- ◇ dfam_version TINYTEXT
- ◇ build_date DATETIME
- ◇ nhmmer_version VARCHAR(45)

mask   TRF Results On Genomic Sequences (not benchmark)
- ◈ seq_accession INT(10)
- ◇ seq_start INT(10)
- ◇ seq_end INT(10)
- ◈ repeat_str VARCHAR(5)
- ◈ repeat_length INT(1)

Indexes
fk_mask_sequence1

hmm_benchmark_region
- ◈ seq_accession INT(10)
- ◇ family_accession VARCHAR(20)
- ◇ seq_start BIGINT(19)
- ◇ seq_end BIGINT(19)
- ◇ strand ENUM('+','-')
- ◇ ali_start BIGINT(19)
- ◇ ali_end BIGINT(19)
- ◇ model_start MEDIUMINT(8)
- ◇ model_end MEDIUMINT(8)
- ◇ hit_bit_score DOUBLE(8,2)
- ◇ hit_evalue_score VARCHAR(15)

Indexes
dfam_reg_region
fk_model_region_dfamseq1

coverage_data
- 🔑 family_accession VARCHAR(20)
- ◇ reversed LONGBLOB
- ◇ forward LONGBLOB
- ◇ nrph LONGBLOB
- ◇ num_full INT(10)
- ◇ num_full_nrph INT(10)
- ◇ num_rev INT(6)
- ◇ karyotype LONGBLOB

Indexes
PRIMARY

model_file
- 🔑 family_accession VARCHAR(20)
- ◇ hit_list LONGBLOB
- ◇ nrph_hit_list LONGBLOB

Indexes
PRIMARY

sequence
- 🔑 accession INT(10)
- ◈ id VARCHAR(80)
- ◇ description TEXT
- ◇ length INT(10)
- ◇ updated TIMESTAMP
- ◇ created DATETIME
- ◈ is_genomic TINYINT(1)

Indexes
PRIMARY
id_UNIQUE
accession_UNIQUE
dfamseq_id
is_genomic_idx

hmm_full_region
- ◈ seq_accession INT(10)
- ◇ family_accession VARCHAR(20)
- ◇ seq_start BIGINT(19)
- ◇ seq_end BIGINT(19)
- ◇ strand ENUM('+','-')
- ◇ ali_start BIGINT(19)
- ◇ ali_end BIGINT(19)
- ◇ model_start MEDIUMINT(8)
- ◇ model_end MEDIUMINT(8)
- ◇ hit_bit_score DOUBLE
- ◇ hit_evalue_score VARCHAR(15)
- ◇ nrph_hit TINYINT(1)
- ◇ divergence DOUBLE

Indexes
hit_bit_score_idx
strand_idx
range_no_acc_idx
ali_start_idx
ali_end_idx
accession

Non-Redundant Coverage, Conservation, and Inserts Graph Data

percentage_id
- 🔑 family_accession VARCHAR(20)
- 🔑 threshold ENUM(...)
- ◇ graph_json LONGBLOB
- ◇ max_insert INT(10)
- ◇ num_seqs INT(10)

Indexes
PRIMARY