STRING v10.0 Database Schema

items
- proteins_names
  - protein_id
  - protein_name
  - species_id
  - source
  - is_preferred_name
- proteins_linkouts
  - protein_id
  - linkout_type
  - is_primary_linkout
- proteins_sequences
  - protein_id
  - sequence
- protein_image
  - protein_id
  - image_id
  - identity
  - source
- protein_orthgroups
  - protein_id
  - orthgroup_id
  - protein_external_id
  - start_position
  - end_position
  - annotation
- proteins_orthgroups
  - orthgroup_id
  - protein_id
- species
  - species_id
  - official_name
  - compact_name
  - kingdom
- species_names
  - species_id
  - species_name
  - official_name
  - is_string_species

network
- node_node_links
  - node_id_a
  - node_id_b
  - combined_score
  - evidence_scores
- best_combined_scores_orthgroups
  - orthgroup_id
  - best_score
- best_combined_scores_proteins
  - protein_id
  - best_score

evidence
- fusion_evidence
  - target_protein_id_a
  - target_protein_id_b
  - source
  - protein
  - species
  - transfer_score
  - fusion_score
- evidence_transfers
  - target_protein_id_a
  - target_protein_id_b
  - source
  - protein
  - species

homology
- best_hit_per_species
  - protein_id
  - species_id
  - nr_high_scoring_hits
- blast_data
  - protein_id
  - species_id
  - score

comments
- not all connections are shown here (for readability).
- the database is designed for speed: no constraints, or triggers - plus, tables are heavily de-normalized (i.e. redundant).
- the tables are served by PostgreSQL (currently in version 9.2.4).

 STRING is locus-based: only a single translated protein per locus is stored (usually the splice-form with the longest open reading frame).

 a) internal identifiers in STRING are simple numericals; they usually remain stable until STRING’s major version number changes.
 b) species identifiers are numericals as well, but they are external and refer to the taxon-identifiers at NCBI.

 c) external protein identifiers:
   - "83333.b1261" taxon-id (species)
   - locus/protein accession (RefSeq, or Ensembl)

d) orthologous groups:
   - “COG0159”
     COG: original groupings (E. Koonin, NCBI)
     KOG: "eukaryotic orthologous group" (NCBI)
     NOG: "non-supervised orthologous group" (STRING/eggNOG)

 e) external gene identifiers:
   - “83333.b1261.NC_000913.1315246”
     accession
     chromosome
     position