

Ensembl Core - Schema documentation

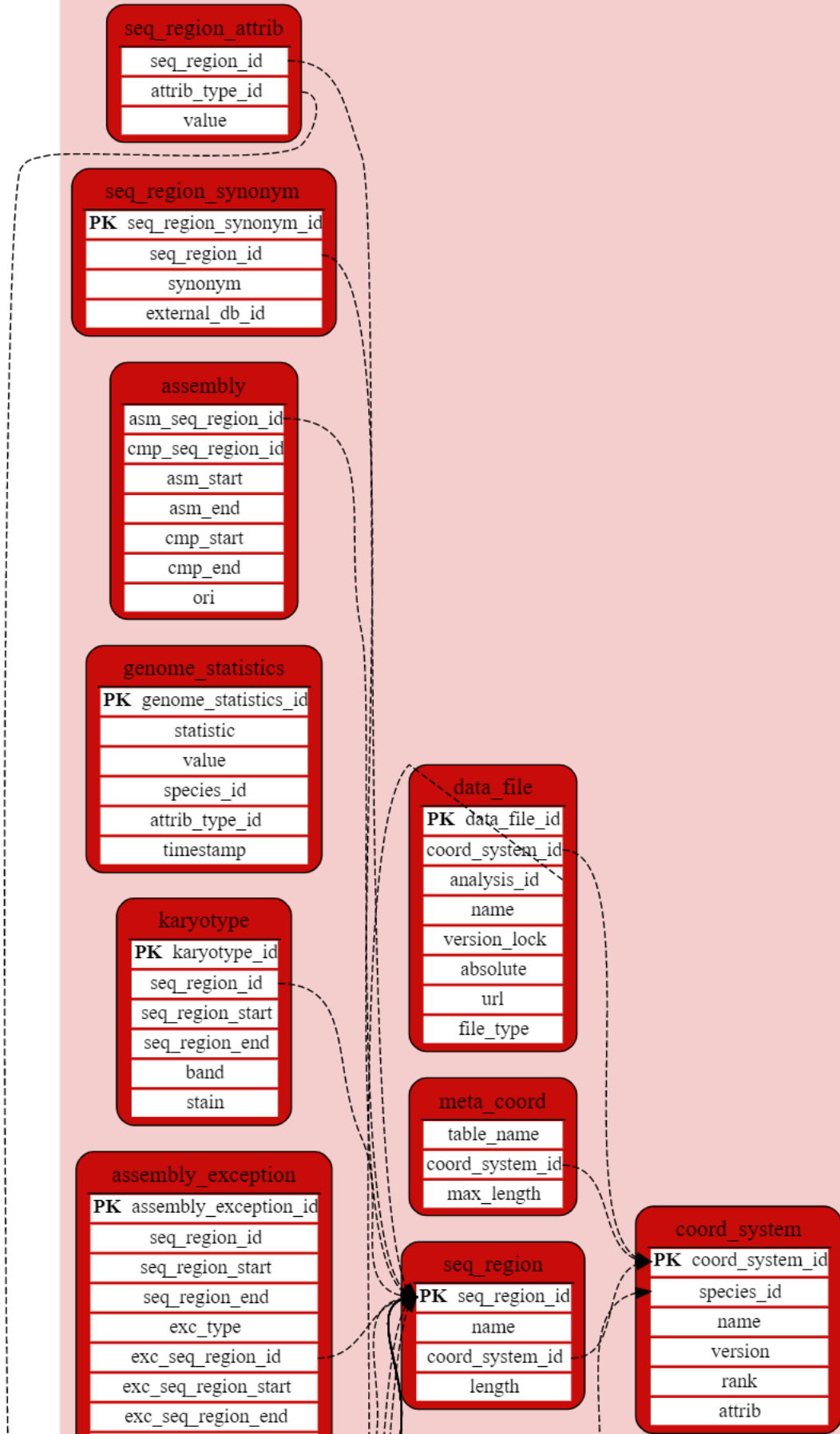
 uswest.ensembl.org/info/docs/api/core/core_schema.html

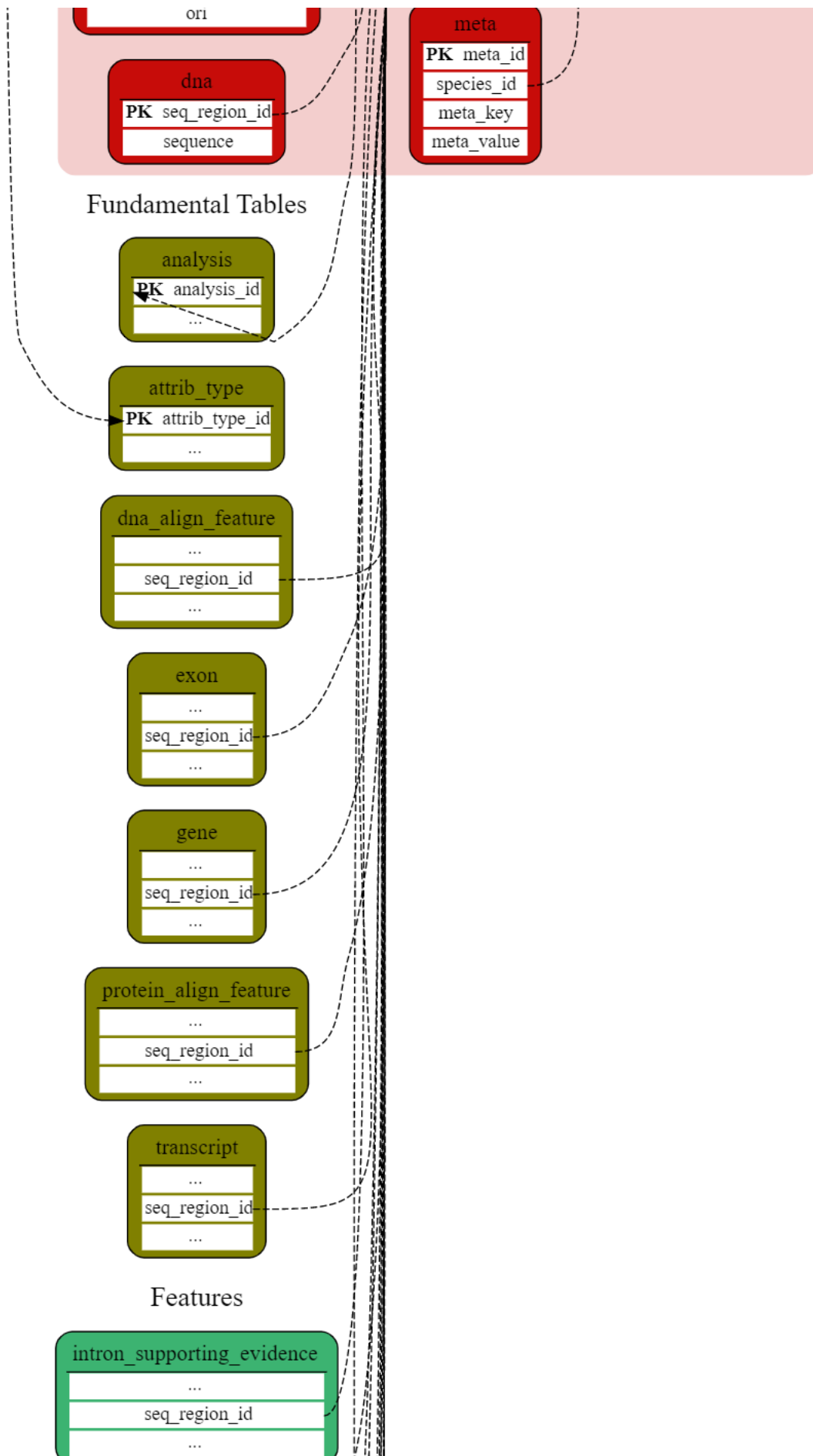
This document gives a high-level description of the tables that make up the Ensembl core schema. Tables are grouped into logical groups, and the purpose of each table is explained. It is intended to allow people to familiarise themselves with the schema when encountering it for the first time, or when they need to use some tables that they've not used before.

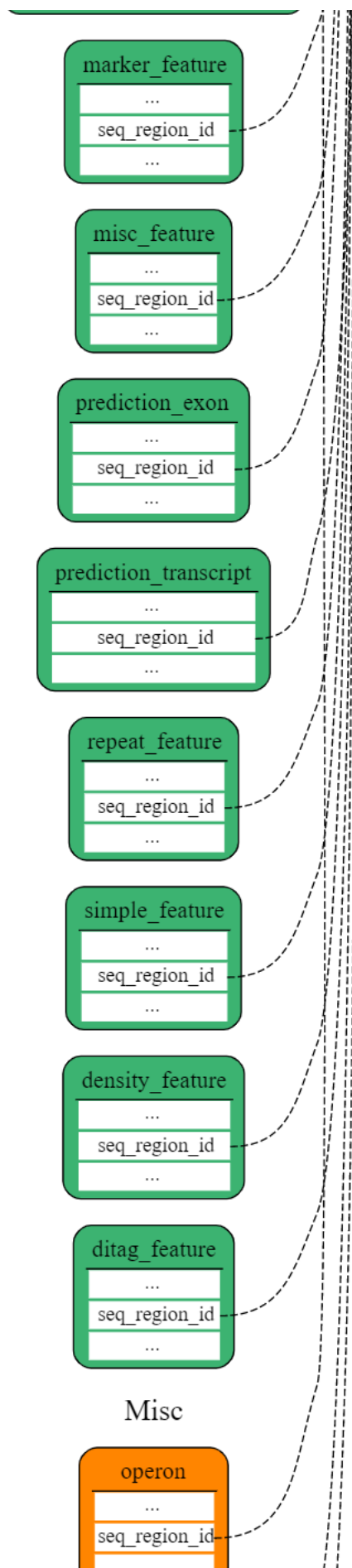
This document refers to version 105 of the Ensembl core schema.

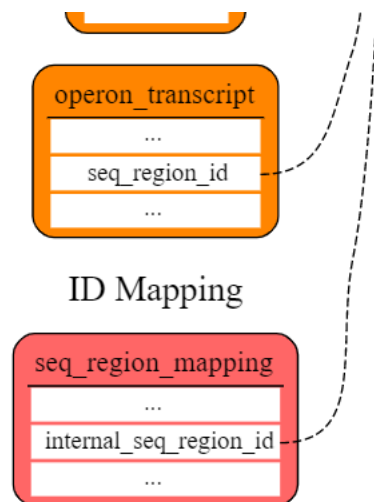
Assembly Tables

Assembly Tables









Core schema diagram: Assembly Tables tables

The assembly table states, which parts of **seq_regions** are exactly equal. It enables to transform coordinates between **seq_regions**. Typically this contains how chromosomes are made of contigs, clones out of contigs, and chromosomes out of supercontigs. It allows you to artificially chunk chromosome sequence into smaller parts. The data in this table defines the "static golden path", i.e. the best effort draft full genome sequence as determined by the UCSC or NCBI (depending which assembly you are using). Each row represents a component, e.g. a contig, (**comp_seq_region_id**, FK from **seq_region** table) at least part of which is present in the golden path. The part of the component that is in the path is delimited by fields **cmp_start** and **cmp_end** (start < end), and the absolute position within the golden path chromosome (or other appropriate assembled structure) (**asm_seq_region_id**) is given by **asm_start** and **asm_end**.

Column	Type	Default value	Description	Index
asm_seq_region_id	INT(10)	-	Assembly sequence region id. Primary key, internal identifier. Foreign key references to the seq_region table.	key: <i>asm_seq_region_idx</i> unique key: <i>all_idx</i>
cmp_seq_region_id	INT(10)	-	Component sequence region id. Foreign key references to the seq_region table.	key: <i>cmp_seq_region_idx</i> unique key: <i>all_idx</i>
asm_start	INT(10)	-	Start absolute position within the golden path chromosome.	key: <i>asm_seq_region_idx</i> unique key: <i>all_idx</i>
asm_end	INT(10)	-	End absolute position within the golden path chromosome.	unique key: <i>all_idx</i>
cmp_start	INT(10)	-	Component start position within the golden path chromosome.	unique key: <i>all_idx</i>
cmp_end	INT(10)	-	Component start position within the golden path chromosome.	unique key: <i>all_idx</i>
ori	TINYINT	-	Orientation: 1 - sense; -1 - antisense.	unique key: <i>all_idx</i>

See also: List of species with populated data: [+ Show species](#)

- [seq_region](#)
- [supercontigs](#)

Allows multiple sequence regions to point to the same sequence, analogous to a symbolic link in a filesystem pointing to the actual file. This mechanism has been implemented specifically to support haplotypes and PARs, but may be useful for other similar structures in the future.

Column	Type	Default value	Description	Index
assembly_exception_id	INT(10)	-	Assembly exception sequence region id. Primary key, internal identifier.	primary key
seq_region_id	INT(10)	-	Sequence region id. Foreign key references to the seq_region table.	key: <i>sr_idx</i>

seq_region_start	INT(10)	-	Sequence start position.	key: <i>sr_idx</i>
seq_region_end	INT(10)	-	Sequence end position.	
exc_type	ENUM: • HAP • PAR • PATCH_FIX • PATCH_NOVEL	NULL	Exception type, e.g. PAR, HAP - haplotype.	
exc_seq_region_id	INT(10)	-	Exception sequence region id. Foreign key references to the seq_region table.	key: <i>ex_idx</i>
exc_seq_region_start	INT(10)	-	Exception sequence start position.	key: <i>ex_idx</i>
exc_seq_region_end	INT(10)	-	Exception sequence end position.	
ori	INT	-	Orientation: 1 - sense; -1 - antisense.	

See also: [List of species with populated data: + Show species](#)

- [assembly](#)
- [seq_region](#)

Stores information about the available co-ordinate systems for the species identified through the `species_id` field. Note that for each species, there must be one co-ordinate system that has the attribute "top_level" and one that has the attribute "sequence_level".

Column	Type	Default value	Description	Index
coord_system_id	INT(10)	-	Primary key, internal identifier.	primary key
species_id	INT(10)	1	Identifies the species for multi-species databases.	unique key: <i>rank_idx</i> unique key: <i>name_idx</i> key: <i>species_idx</i>
name	VARCHAR(40)	-	Co-ordinate system name, e.g. 'chromosome', 'contig', 'scaffold' etc.	unique key: <i>name_idx</i>
version	VARCHAR(255)	NULL	Assembly.	unique key: <i>name_idx</i>
rank	INT	-	Co-ordinate system rank.	unique key: <i>rank_idx</i>
attrib	SET: • default_version • sequence_level	NULL	Co-ordinate system attrib (e.g. "top_level", "sequence_level").	

See also: [List of species with populated data: + Show species](#)

- [seq_region](#)
- [meta_coord](#)
- [meta](#)

Allows the storage of flat file locations used to store large quantities of data currently unsuitable in a traditional database table.

Column	Type	Default value	Description	Index
data_file_id	INT(10)	-	Auto-increment surrogate primary key	primary key
coord_system_id	INT(10)	-	Coordinate system this file is linked to. Used to decipher the assembly version it was mapped to	unique key: <i>df_unq_idx</i>

analysis_id	SMALLINT	-	Analysis this file is linked to	unique key: <i>df_unq_idx</i> key: <i>df_analysis_idx</i>
name	VARCHAR(100)	-	Name of the file	unique key: <i>df_unq_idx</i> key: <i>df_name_idx</i>
version_lock	TINYINT(1)	0	Indicates that this file is only compatible with the current Ensembl release version	
absolute	TINYINT(1)	0	Flags that the URL given is fully resolved and should be used without question	
url	TEXT	NULL	Optional path to the file (can be absolute or relative)	
file_type	ENUM: • BAM • BAMCOV • BIGBED • BIGWIG • VCF	NULL	Type of file e.g. BAM, BIGBED, BIGWIG and VCF	unique key: <i>df_unq_idx</i>

Contains DNA sequence. This table has a 1:1 relationship with the `seq_region` table.

Column	Type	Default value	Description	Index
seq_region_id	INT(10)	-	Primary key, internal identifier. Foreign key references to the <code>seq_region</code> table.	primary key
sequence	LONGTEXT	-	DNA sequence.	

See also: [List of species with populated data: !\[\]\(feabb98897b440bc8695a03336a6e2df_img.jpg\) Show species](#)
[seq_region](#)

Contains genome and assembly related statistics These include but are not limited to: feature counts, sequence lengths

Column	Type	Default value	Description	Index
genome_statistics_id	INT(10)	-	Primary key, internal identifier.	primary key
statistic	VARCHAR(128)	-	Name of the statistics	unique key: <i>stats_uniq</i>
value	BIGINT(11)	'0'	Corresponding value of the statistics (count/length)	
species_id	INT	1	Identifies the species for multi-species databases.	unique key: <i>stats_uniq</i>
attrib_type_id	INT(10)	NULL	To distinguish similar statistics for different cases	unique key: <i>stats_uniq</i>
timestamp	DATETIME	NULL	Date the statistics was generated	

List of species with populated data: [+ Show species](#)

Describes bands that can be stained on the chromosome.

Column	Type	Default value	Description	Index
karyotype_id	INT(10)	-	Primary key, internal identifier.	primary key
seq_region_id	INT(10)	-	Foreign key references to the <code>seq_region</code> table.	key: <i>region_band_idx</i>
seq_region_start	INT(10)	-	Sequence start position.	
seq_region_end	INT(10)	-	Sequence end position.	
band	VARCHAR(40)	NULL	Band.	key: <i>region_band_idx</i>
stain	VARCHAR(40)	NULL	Stain.	

See also: [List of species with populated data: + Show species](#)

[seq_region](#)

Stores data about the data in the current schema. Taxonomy information, version information and the default value for the type column in the assembly table are stored here. Unlike other tables, data in the meta table is stored as key-value pairs. Also stores (via assembly.mapping keys) the relationships between co-ordinate systems in the assembly table. The species_id field of the meta table is used in multi-species databases and makes it possible to have species-specific meta key-value pairs. The species-specific meta key-value pairs needs to be repeated for each species_id. Entries in the meta table that are not specific to any one species, such as the schema_version key and any other schema-related information must have their species_id field set to NULL. The default species_id, and the only species_id value allowed in single-species databases, is 1.

Column	Type	Default value	Description	Index
meta_id	INT	-	Primary key, internal identifier.	primary key
species_id	INT	1	Identifies the species for multi-species databases.	unique key: species_key_value_idx key: species_value_idx
meta_key	VARCHAR(40)	-	Name of the meta entry, e.g. "schema_version".	unique key: species_key_value_idx
meta_value	VARCHAR(255)	-	Corresponding value of the key, e.g. "61".	unique key: species_key_value_idx key: species_value_idx

See also: [List of species with populated data: + Show species](#)

- [assembly](#)
- [coord_system](#)

Describes which co-ordinate systems the different feature tables use.

Column	Type	Default value	Description	Index
table_name	VARCHAR(40)	-	Ensembl database table name.	unique key: cs_table_name_idx
coord_system_id	INT(10)	-	Foreign key references to the coord_system table.	unique key: cs_table_name_idx
max_length	INT	NULL	Longest sequence length.	

See also: [List of species with populated data: + Show species](#)

[coord_system](#)

Stores information about sequence regions. The primary key is used as a pointer into the dna table so that actual sequence can be obtained, and the coord_system_id allows sequence regions of multiple types to be stored. Clones, contigs and chromosomes are all now stored in the seq_region table. Contigs are stored with the co-ordinate system 'contig'. The relationship between contigs and clones is stored in the assembly table. The relationships between contigs and chromosomes, and between contigs and supercontigs, are stored in the assembly table.

Column	Type	Default value	Description	Index
seq_region_id	INT(10)	-	Primary key, internal identifier.	primary key
name	VARCHAR(255)	-	Sequence region name.	unique key: name_cs_idx
coord_system_id	INT(10)	-	Foreign key references to the coord_system table.	unique key: name_cs_idx key: cs_idx
length	INT(10)	-	Sequence length.	

See also: [List of species with populated data: + Show species](#)

- [dna](#)
- [coord_system](#)

Allows "attributes" to be defined for certain seq_regions. Provides a way of storing extra information about particular seq_regions without adding extra columns to the seq_region table. e.g.

Column	Type	Default value	Description	Index
seq_region_id	INT(10)	'0'	Foreign key references to the seq_region table.	key: <i>seq_region_idx</i> unique key: <i>region_attribx</i>
attrib_type_id	SMALLINT(5)	'0'	Foreign key references to the attrib_type table.	key: <i>type_val_idx</i> unique key: <i>region_attribx</i>
value	TEXT	-	Attribute value.	key: <i>type_val_idx</i> key: <i>val_only_idx</i> unique key: <i>region_attribx</i>

See also: List of species with populated data: [+ Show species](#)

- [seq_region](#)
- [attrib_type](#)

Allows for storing multiple names for sequence regions.

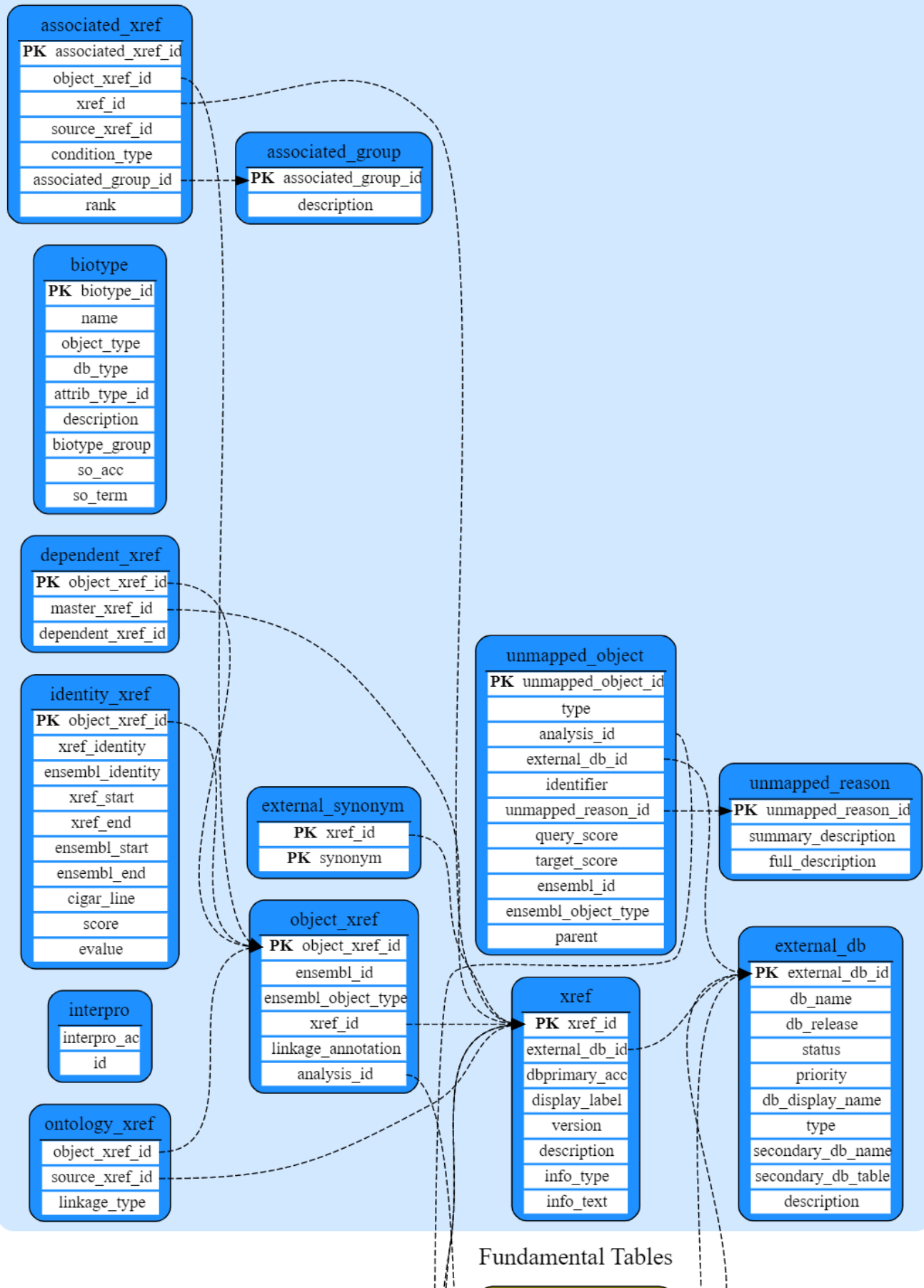
Column	Type	Default value	Description	Index
seq_region_synonym_id	INT	-	Primary key, internal identifier.	primary key
seq_region_id	INT(10)	-	Foreign key references to the seq_region table.	unique key: <i>syn_idx</i> key: <i>seq_region_idx</i>
synonym	VARCHAR(250)	-	Alternative name for sequence region.	unique key: <i>syn_idx</i>
external_db_id	INT	NULL	Foreign key references to the external_db table.	

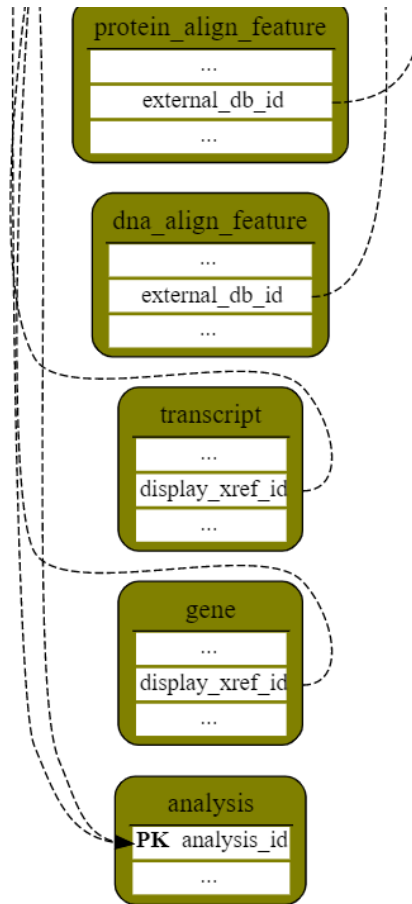
See also: List of species with populated data: [+ Show species](#)

- [seq_region](#)
- [external_db](#)

External References

External References





Core schema diagram: External References tables

Groups together xref associations under a single description. Used when more than one associated xref term must be used to describe a condition

Column	Type	Default value	Description	Index
associated_group_id	INT(10)	-	Associated group id. Primary key, internal identifier	primary key
description	VARCHAR(128)	NULL	Optional description for this group	

See also: [List of species with populated data: !\[\]\(cbe80b694ebd74fcfe136a095b608235_img.jpg\) Show species](#)

[associated_xref](#)

This table associates extra associated annotations with a given ontology xref evidence and source under a specific condition. For GO this allows qualifiers (with/from) or annotation extensions to be added to a given ontology annotation.

Column	Type	Default value	Description	Index
associated_xref_id	INT(10)	-	Associated xref id. Primary key, internal identifier	primary key
object_xref_id	INT(10)	'0'	Object xref id this associated xref is linked to. Foreign key linked to the object_xref table	key: <i>associated_object_idx</i> unique key: <i>object_associated_source_type_idx</i>
xref_id	INT(10)	'0'	Xref which is the associated term. Foreign key linked to the xref table	key: <i>associated_idx</i> unique key: <i>object_associated_source_type_idx</i>
source_xref_id	INT(10)	NULL	Xref which is source of this association. Foreign key linked to the xref table	key: <i>associated_source_idx</i> unique key: <i>object_associated_source_type_idx</i>

condition_type	VARCHAR(128)	NULL	The type of condition this link occurs in e.g. evidence, from, residue or assigned_by	unique key: <i>object_associated_source_type_idx</i>
associated_group_id	INT(10)	NULL	Foreign key to allow for <u>associated_group</u>	key: <i>associated_group_idx</i> unique key: <i>object_associated_source_type_idx</i>
rank	INT(10)	'0'	The rank in which the association occurs within an <u>associated_group</u>	

See also: List of species with populated data: [+ Show species](#)

- [object_xref](#)
- [associated_group](#)
- [xref](#)

Stores data about the biotypes and mappings to Sequence Ontology.

Column	Type	Default value	Description	Index
biotype_id	INT	-	Primary key, internal identifier.	primary key
name	VARCHAR(64)	-	Ensembl biotype name.	unique key: <i>name_type_idx</i>
object_type	ENUM: • gene • transcript	'gene'	Ensembl object type: 'gene' or 'transcript'.	unique key: <i>name_type_idx</i>
db_type	SET: • cdna • core • coreexpressionatlas • coreexpressionest • coreexpressiongnf • funcgen • otherfeatures • maseq • variation • vega • presite • sangervega	'core'	Type, e.g. 'cdna', 'core', 'coreexpressionatlas', 'coreexpressionest', 'coreexpressiongnf', 'funcgen', 'otherfeatures', 'maseq', 'variation', 'vega', 'presite', 'sangervega'	
attrib_type_id	INT	NULL	Foreign key references to the <u>attrib_type</u> table.	
description	TEXT	NULL	Description.	
biotype_group	ENUM: • coding • pseudogene • snoncoding • lnoncoding • mnoncoding • LRG • undefined • no_group	NULL	Group, e.g. 'coding', 'pseudogene', 'snoncoding', 'lnoncoding', 'mnoncoding', 'LRG', 'undefined', 'no_group'	
so_acc	VARCHAR(64)	NULL	Sequence Ontology accession of the biotype.	
so_term	VARCHAR(1023)	NULL	Sequence Ontology term of the biotype.	

See also: List of species with populated data: [+ Show species](#)

[attrib_type](#)

Describes dependent external references which can't be directly mapped to Ensembl entities. They are linked to primary external references instead.

Column	Type	Default value	Description	Index
object_xref_id	INT(10)	-	Primary key, internal identifier. Foreign key references to the <u>object_xref</u> table.	primary key
master_xref_id	INT(10)	-	Foreign key references to the <u>xref</u> table.	key: <i>master_idx</i>
dependent_xref_id	INT(10)	-	Foreign key references to the <u>xref</u> table.	key: <i>dependent</i>

See also: List of species with populated data: [+ Show species](#)

- [xref](#)
- [object_xref](#)

Stores data about the external databases in which the objects described in the xref table are stored.

Column	Type	Default value	Description	Index
external_db_id	INT	-	Primary key, internal identifier.	primary key
db_name	VARCHAR(100)	-	Database name.	unique key: <i>db_name_o</i>
db_release	VARCHAR(255)	NULL	Database release.	unique key: <i>db_name_o</i>
status	ENUM: • KNOWNXREF • KNOWN • XREF • PRED • ORTH • PSEUDO	-	Status, e.g. 'KNOWNXREF', 'KNOWN', 'XREF', 'PRED', 'ORTH', 'PSEUDO'.	
priority	INT	-	Determines which one of the xrefs will be used as the gene name.	
db_display_name	VARCHAR(255)	NULL	Database display name.	
type	ENUM: • ARRAY • ALT_TRANS • ALT_GENE • MISC • LIT • PRIMARY_DB_SYNONYM • ENSEMBL	-	Type, e.g. 'ARRAY', 'ALT_TRANS', 'ALT_GENE', 'MISC', 'LIT', 'PRIMARY_DB_SYNONYM', 'ENSEMBL'.	
secondary_db_name	VARCHAR(255)	NULL	Secondary database name.	
secondary_db_table	VARCHAR(255)	NULL	Secondary database table.	
description	TEXT	NULL	Description.	

See also: List of species with populated data: [+ Show species](#)

- [xref](#)
- [unmapped_object](#)
- [protein_align_feature](#)
- [dna_align_feature](#)

Some xref objects can be referred to by more than one name. This table relates names to xref IDs.

Column	Type	Default value	Description	Index
xref_id	INT(10)	-	Primary key, internal identifier.	primary key

synonym	VARCHAR(100)	-	Synonym	primary key key: <i>name_index</i>
----------------	--------------	---	---------	---------------------------------------

See also: List of species with populated data: [+ Show species](#)
[xref](#)

Describes how well a particular xref object matches the Ensembl object.

Column	Type	Default value	Description	Index
object_xref_id	INT(10)	-	Primary key, internal identifier. Foreign key references to the object_xref table.	primary key
xref_identity	INT(5)	NULL	Percentage identity.	
ensembl_identity	INT(5)	NULL	Percentage identity.	
xref_start	INT	NULL	Xref sequence start.	
xref_end	INT	NULL	Xref sequence end.	
ensembl_start	INT	NULL	Ensembl sequence start.	
ensembl_end	INT	NULL	Ensembl sequence end.	
cigar_line	TEXT	NULL	Used to encode gapped alignments.	
score	DOUBLE	NULL	Match score.	
evalue	DOUBLE	NULL	Match evalue.	

See also: List of species with populated data: [+ Show species](#)
[object_xref](#)

Allows storage of links to the InterPro database. InterPro is a database of protein families, domains and functional sites in which identifiable features found in known proteins can be applied to unknown protein sequences. [InterPro](#) - The InterPro website

Column	Type	Default value	Description	Index
interpro_ac	VARCHAR(40)	-	InterPro protein accession number.	unique key: <i>accession_idx</i>
id	VARCHAR(40)	-	InterPro protein id.	unique key: <i>accession_idx</i> key: <i>id_idx</i>

List of species with populated data: [+ Show species](#)

Describes links between Ensembl objects and objects held in external databases. The Ensembl object can be one of several types; the type is held in the `ensembl_object_type` column. The ID of the particular Ensembl gene, translation or whatever is given in the `ensembl_id` column. The `xref_id` points to the entry in the `xref` table that holds data about the external object. Each Ensembl object can be associated with zero or more xrefs. An xref object can be associated with one or more Ensembl objects.

Column	Type	Default value	Description	Index
object_xref_id	INT(10)	-	Primary key, internal identifier.	primary key
ensembl_id	INT(10)	-	Foreign key references to the seq_region , transcript , gene , @translation tables depending on <code>ensembl_object_type</code> .	unique key: <i>xref_idx</i> key: <i>ensembl_idx</i>

ensembl_object_type	ENUM: <ul style="list-style-type: none"> • RawContig • Transcript • Gene • Translation • Operon • OperonTranscript • Marker • RNAProduct 	NULL	Ensembl object type: 'RawContig', 'Transcript', 'Gene', 'Translation', ..., 'RNAProduct'	unique key: <i>xref_idx</i> key: <i>ensembl_idx</i>
xref_id	INT(10)	-	Foreign key references to the <u>xref</u> table.	unique key: <i>xref_idx</i>
linkage_annotation	VARCHAR(255)	NULL	Additional annotation on the linkage.	
analysis_id	SMALLINT	NULL	Foreign key references to the <u>analysis</u> table.	unique key: <i>xref_idx</i> key: <i>analysis_idx</i>

See also: List of species with populated data: [+ Show species](#)

- [xref](#)
- [identity_xref](#)
- [seq_region](#)
- [transcript](#)
- [gene](#)
- [analysis](#)

This table associates Evidence Tags to the relationship between EnsEMBL objects and ontology accessions (primarily GO accessions). The relationship to GO that is stored in the database is actually derived through the relationship of EnsEMBL peptides to SwissProt peptides, i.e. the relationship is derived like this: ENSP -> SWISSPROT -> GO And the evidence tag describes the relationship between the SwissProt Peptide and the GO entry. In reality, however, we store this in the database like this: ENSP -> SWISSPROT ENSP -> GO and the evidence tag hangs off of the relationship between the ENSP and the GO identifier. Some ENSPs are associated with multiple closely related Swissprot entries which may both be associated with the same GO identifier but with different evidence tags. For this reason a single Ensembl - external db object relationship in the object_xref table can be associated with multiple evidence tags in the ontology_xref table.

Column	Type	Default value	Description	Index
object_xref_id	INT(10)	'0'	Composite key. Foreign key references to the <u>object_xref</u> table.	key: <i>object_idx</i> unique key: <i>object_source_type_idx</i>
source_xref_id	INT(10)	NULL	Composite key. Foreign key references to the <u>xref</u> table.	key: <i>source_idx</i> unique key: <i>object_source_type_idx</i>
linkage_type	VARCHAR(3)	NULL	Composite key. <u>Evidence tags</u>	unique key: <i>object_source_type_idx</i>

See also: List of species with populated data: [+ Show species](#)

- [object_xref](#)
- [xref](#)

Describes why a particular external entity was not mapped to an ensembl one.

Column	Type	Default value	Description	Index
unmapped_object_id	INT(10)	-	Primary key, internal identifier.	primary key
type	ENUM: <ul style="list-style-type: none"> • xref • cDNA • Marker 	-	Object type: 'xref', 'cDNA', 'Marker'.	
analysis_id	SMALLINT	-	Foreign key references to the <u>analysis</u> table.	key: <i>anal_exdb_idx</i>

external_db_id	INT	NULL	Foreign key references to the <u>external_db</u> table.	unique key: <i>unique_unmapped_obj_idx</i> key: <i>anal_exdb_idx</i> key: <i>ext_db_identifier_idx</i>
identifier	VARCHAR(255)	-	External database identifier.	unique key: <i>unique_unmapped_obj_idx</i> key: <i>id_idx</i> key: <i>ext_db_identifier_idx</i>
unmapped_reason_id	INT(10)	-	Foreign key references to the <u>unmapped_reason</u> table.	unique key: <i>unique_unmapped_obj_idx</i>
query_score	DOUBLE	NULL	Actual mapping query score.	
target_score	DOUBLE	NULL	Target mapping query score.	
ensembl_id	INT(10)	'0'	Foreign key references to the <u>seq_region</u> , <u>transcript</u> , <u>gene</u> , @translation tables depending on <u>ensembl_object_type</u> .	unique key: <i>unique_unmapped_obj_idx</i>
ensembl_object_type	ENUM: • RawContig • Transcript • Gene • Translation	'RawContig'	Ensembl object type: 'RawContig', 'Transcript', 'Gene', 'Translation'.	unique key: <i>unique_unmapped_obj_idx</i>
parent	VARCHAR(255)	NULL	Foreign key references to the <u>dependent_xref</u> table, in case the unmapped object is dependent on a primary external reference which wasn't mapped to an ensembl one.	unique key: <i>unique_unmapped_obj_idx</i>

See also: List of species with populated data: [+ Show species](#)

- [analysis](#)
- [external_db](#)
- [unmapped_reason](#)
- [seq_region](#)
- [transcript](#)
- [gene](#)
- [dependent_xref](#)

Describes the reason why a mapping failed.

Column	Type	Default value	Description	Index
unmapped_reason_id	INT(10)	-	Primary key, internal identifier.	primary key
summary_description	VARCHAR(255)	NULL	Summarised description.	
full_description	VARCHAR(255)	NULL	Full description.	

See also: List of species with populated data: [+ Show species](#)

[unmapped_object](#)

Holds data about objects which are external to Ensembl, but need to be associated with Ensembl objects. Information about the database that the external object is stored in is held in the external_db table entry referred to by the external_db column.

Column	Type	Default value	Description	Index
xref_id	INT(10)	-	Primary key, internal identifier.	primary key
external_db_id	INT	-	Foreign key references to the <u>external_db</u> table.	unique key: <i>id_index</i>
dbprimary_acc	VARCHAR(512)	-	Primary accession number.	unique key: <i>id_index</i>
display_label	VARCHAR(512)	-	Display label for the Ensembl web site.	key: <i>display_index</i>

version	VARCHAR(10)	NULL	Object version.	unique key: <i>id_index</i>
description	TEXT	NULL	Object description.	
info_type	ENUM: <ul style="list-style-type: none"> NONE PROJECTION MISC DEPENDENT DIRECT SEQUENCE_MATCH INFERRED_PAIR PROBE UNMAPPED COORDINATE_OVERLAP CHECKSUM 	'NONE'	'PROJECTION', 'MISC', 'DEPENDENT', 'DIRECT', 'SEQUENCE_MATCH', 'INFERRED_PAIR', 'PROBE', 'UNMAPPED', 'COORDINATE_OVERLAP', 'CHECKSUM'.	unique key: <i>id_index</i> key: <i>info_type_idx</i>
info_text	VARCHAR(255)	"	Text	unique key: <i>id_index</i>

See also: List of species with populated data: [+ Show species](#)

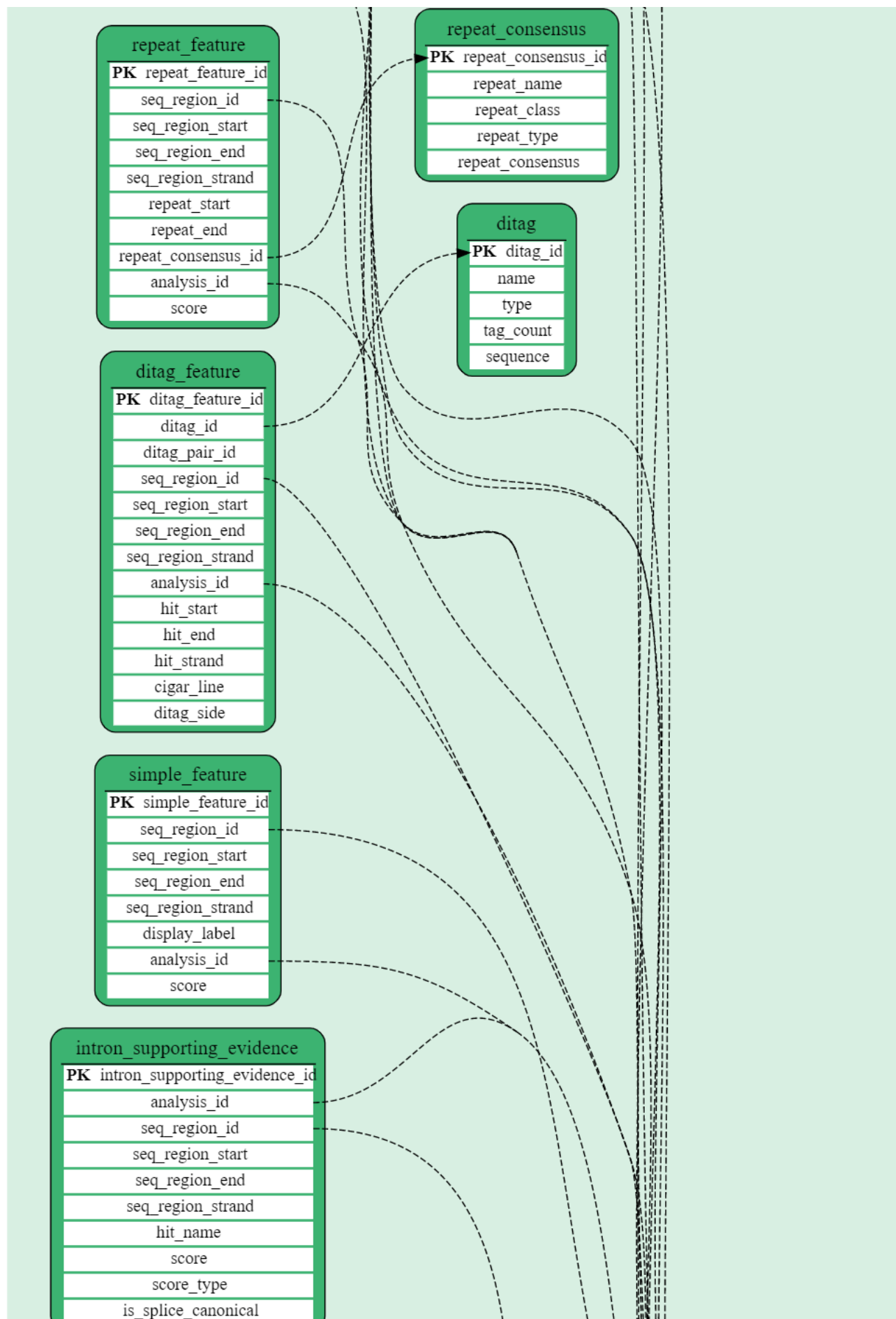
- [associated_xref](#)
- [dependent_xref](#)
- [gene](#)
- [transcript](#)
- [external_db](#)
- [external_synonym](#)
- [object_xref](#)
- [ontology_xref](#)

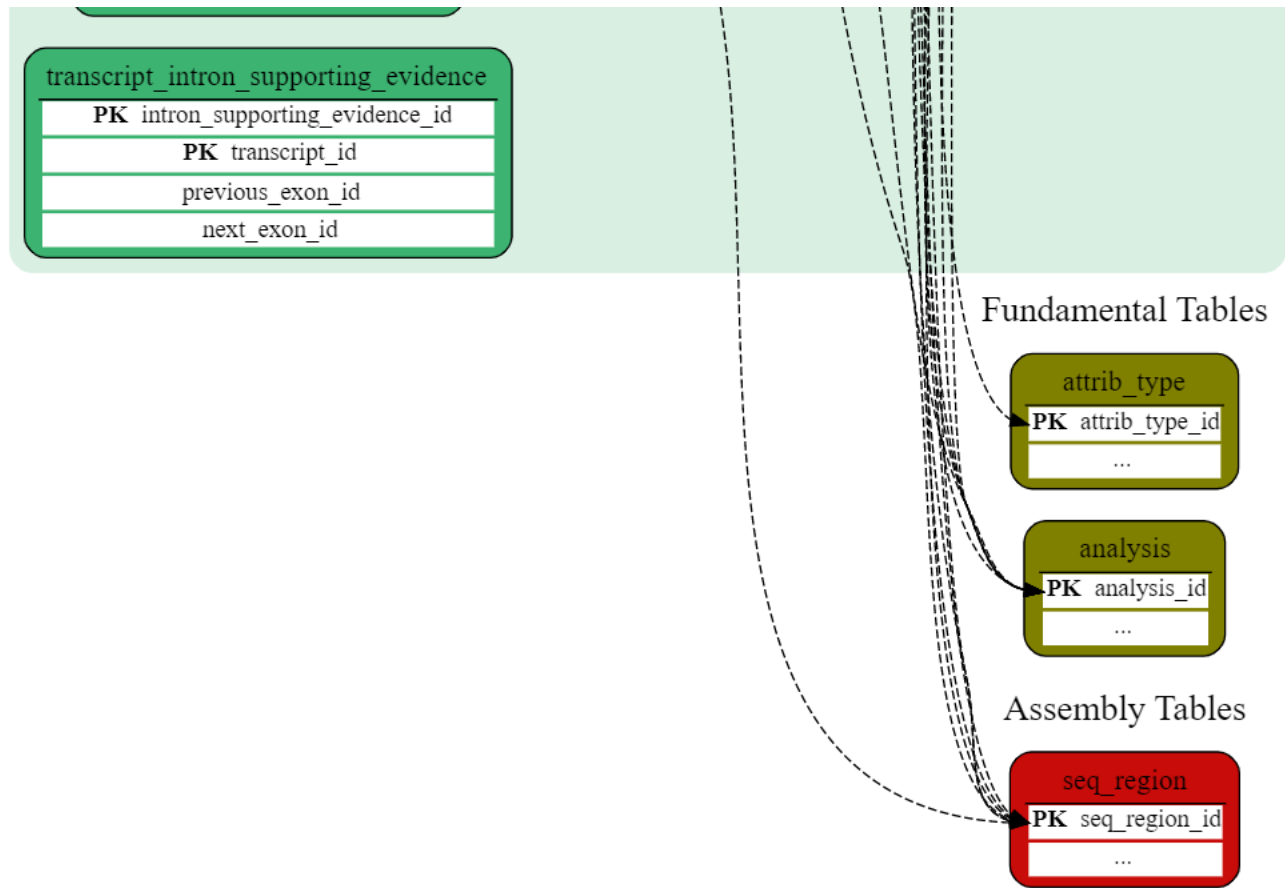
Features

Features

The diagram illustrates the relationships between 12 database tables. Each table is represented by a green box containing its name and a list of attributes, with primary keys (PK) highlighted. Dashed arrows indicate foreign key relationships from a table's attribute to another table's primary key.

- marker_map_location** (PK: marker_id, map_id)
 - marker_id → **marker** (PK: marker_id)
 - map_id → **map** (PK: map_id)
- map** (PK: map_id, map_name)
- misc_set** (PK: misc_set_id)
 - misc_set_id → **marker_map_location** (FK: map_id)
 - misc_set_id → **marker_synonym** (FK: marker_id)
 - misc_set_id → **marker** (FK: marker_id)
 - misc_set_id → **marker_feature** (FK: marker_id)
 - misc_set_id → **misc_attrb** (FK: misc_feature_id)
 - misc_set_id → **marker_synonym** (FK: marker_id)
- marker_synonym** (PK: marker_synonym_id, marker_id, source, name)
 - marker_id → **marker** (FK: marker_id)
- marker** (PK: marker_id, display_marker_synonym_id, left_primer, right_primer, min_primer_dist, max_primer_dist, priority, type)
 - display_marker_synonym_id → **marker_synonym** (FK: marker_synonym_id)
- misc_attrb** (misc_feature_id, attrib_type_id, value)
 - misc_feature_id → **marker** (FK: marker_id)
 - misc_feature_id → **marker_feature** (FK: marker_id)
 - misc_feature_id → **misc_feature** (FK: misc_feature_id)
- marker_feature** (PK: marker_feature_id, marker_id, seq_region_id, seq_region_start, seq_region_end, analysis_id, map_weight)
 - marker_id → **marker** (FK: marker_id)
 - seq_region_id → **marker** (FK: marker_id)
 - seq_region_id → **misc_feature** (FK: seq_region_id)
 - seq_region_start → **marker** (FK: marker_id)
 - seq_region_end → **marker** (FK: marker_id)
 - analysis_id → **marker** (FK: marker_id)
 - map_weight → **marker** (FK: marker_id)
- misc_feature** (PK: misc_feature_id, seq_region_id, seq_region_start, seq_region_end, seq_region_strand)
 - seq_region_id → **marker** (FK: marker_id)
 - seq_region_id → **marker_feature** (FK: seq_region_id)
 - seq_region_start → **marker** (FK: marker_id)
 - seq_region_end → **marker** (FK: marker_id)
 - seq_region_strand → **marker** (FK: marker_id)
- density_feature** (PK: density_feature_id, density_type_id, seq_region_id, seq_region_start, seq_region_end, density_value)
 - density_type_id → **density_type** (FK: density_type_id)
 - seq_region_id → **marker** (FK: marker_id)
 - seq_region_id → **marker_feature** (FK: seq_region_id)
 - seq_region_start → **marker** (FK: marker_id)
 - seq_region_end → **marker** (FK: marker_id)
- density_type** (PK: density_type_id, analysis_id, block_size, region_features, value_type)
 - analysis_id → **marker** (FK: marker_id)
 - analysis_id → **marker_feature** (FK: analysis_id)
 - analysis_id → **marker_synonym** (FK: analysis_id)
 - analysis_id → **marker** (FK: marker_id)
 - analysis_id → **marker_feature** (FK: analysis_id)
 - analysis_id → **marker_synonym** (FK: analysis_id)
- prediction_exon** (PK: prediction_exon_id, prediction_transcript_id, exon_rank, seq_region_id, seq_region_start, seq_region_end, seq_region_strand, start_phase, score, p_value)
 - prediction_transcript_id → **prediction_transcript** (FK: prediction_transcript_id)
 - seq_region_id → **marker** (FK: marker_id)
 - seq_region_id → **marker_feature** (FK: seq_region_id)
 - seq_region_start → **marker** (FK: marker_id)
 - seq_region_end → **marker** (FK: marker_id)
 - seq_region_strand → **marker** (FK: marker_id)
- prediction_transcript** (PK: prediction_transcript_id, seq_region_id, seq_region_start, seq_region_end, seq_region_strand, analysis_id, display_label)
 - seq_region_id → **marker** (FK: marker_id)
 - seq_region_id → **marker_feature** (FK: seq_region_id)
 - seq_region_start → **marker** (FK: marker_id)
 - seq_region_end → **marker** (FK: marker_id)
 - seq_region_strand → **marker** (FK: marker_id)
 - analysis_id → **marker** (FK: marker_id)
 - analysis_id → **marker_feature** (FK: analysis_id)
 - analysis_id → **marker_synonym** (FK: analysis_id)





Core schema diagram: Features tables

Describes features representing a density, or precentage coverage etc. in a given region.

Column	Type	Default value	Description	Index
density_feature_id	INT(10)	-	Primary key, internal identifier.	primary key
density_type_id	INT(10)	-	Density type. Foreign key references to the density_type table.	key: seq_region_idx
seq_region_id	INT(10)	-	Sequence region. Foreign key references to the seq_region table.	key: seq_region_idx key: seq_region_id_idx
seq_region_start	INT(10)	-	Sequence start position.	key: seq_region_idx
seq_region_end	INT(10)	-	Sequence end position.	
density_value	FLOAT	-	Density value.	

See also: List of species with populated data: [+ Show species](#)

- [density_type](#)
- [seq_region](#)

Describes type representing a density, or percentage coverage etc. in a given region.

Column	Type	Default value	Description	Index
density_type_id	INT(10)	-	Primary key, internal identifier.	primary key
analysis_id	SMALLINT	-	Foreign key references to the analysis table.	unique key: analysis_idx
block_size	INT	-	Block size.	unique key: analysis_idx

region_features	INT	-	The number of features per sequence region inside this density type.	unique key: <i>analysis_idx</i>
value_type	ENUM: • sum • ratio	-	Value type, e.g. 'sum', 'ratio'.	

See also: List of species with populated data: [+ Show species](#)

- [density_feature](#)
- [analysis](#)

Represents a ditag object in the Ensembl database. Corresponds to original tag containing the full sequence. This can be a single piece of sequence like CAGE tags or a ditag with concatenated sequence from 5' and 3' end like GIS or GSC tags. This data is available as a DAS track in ContigView on the Ensembl web site.

Column	Type	Default value	Description	Index
ditag_id	INT(10)	-	Primary key, internal identifier.	primary key
name	VARCHAR(30)	-	Ditag name.	
type	VARCHAR(30)	-	Ditag type.	
tag_count	smallint(6)	1	Tag count.	
sequence	TINYTEXT	-	Sequence.	

See also:

[ditag_feature](#)

Describes where ditags hit on the genome. Represents a mapped ditag object in the Ensembl database. These are the original tags separated into start ("L") and end ("R") parts if applicable, successfully aligned to the genome. Two DitagFeatures usually relate to one parent Ditag. Alternatively there are CAGE tags e.g. which only have a 5' tag ("F").

Column	Type	Default value	Description	Index
ditag_feature_id	INT(10)	-	Primary key, internal identifier.	primary key
ditag_id	INT(10)	'0'	Foreign key references to the ditag table.	key: <i>ditag_idx</i>
ditag_pair_id	INT(10)	'0'	Ditag pair id.	key: <i>ditag_pair_idx</i>
seq_region_id	INT(10)	'0'	Foreign key references to the seq_region table.	key: <i>seq_region_idx</i>
seq_region_start	INT(10)	'0'	Sequence start position.	key: <i>seq_region_idx</i>
seq_region_end	INT(10)	'0'	Sequence end position.	key: <i>seq_region_idx</i>
seq_region_strand	TINYINT(1)	'0'	Sequence region strand: 1 - forward; -1 - reverse.	
analysis_id	SMALLINT	'0'	Foreign key references to the analysis table.	
hit_start	INT(10)	'0'	Alignment hit start position.	
hit_end	INT(10)	'0'	Alignment hit end position.	
hit_strand	TINYINT(1)	'0'	Alignment hit strand: 1 - forward; -1 - reverse.	
cigar_line	TINYTEXT	-	Used to encode gapped alignments.	
ditag_side	ENUM: • F • L • R	-	Ditag side: L - start, R - end, F - 5'tag only	

See also:

- [ditag](#)
- [seq_region](#)
- [analysis](#)

Provides the evidence which we have used to declare an intronic region

Column	Type	Default value	Description	Index
intron_supporting_evidence_id	INT(10)	-	Surrogate primary key	primary key
analysis_id	SMALLINT	-	Foreign key references to the analysis table.	unique: <i>key</i>
seq_region_id	INT(10)	-	Foreign key references to the seq_region table.	unique: <i>key</i> key: <i>seq_region_idx</i>
seq_region_start	INT(10)	-	Sequence start position.	unique: <i>key</i> key: <i>seq_region_idx</i>
seq_region_end	INT(10)	-	Sequence end position.	unique: <i>key</i>
seq_region_strand	TINYINT(2)	-	Sequence region strand: 1 - forward; -1 - reverse.	unique: <i>key</i>
hit_name	VARCHAR(100)	-	External entity name/identifier.	unique: <i>key</i>
score	DECIMAL(10,3)	NULL	Score supporting the intron	
score_type	ENUM: <ul style="list-style-type: none">• NONE• DEPTH	'NONE'	The type of score e.g. NONE	
is_splice_canonical	TINYINT(1)	0	Indicates if the splice junction can be considered canonical i.e. behaves according to accepted rules	

See also:

List of species with populated data: [+ Show species](#)

- [transcript_intron_supporting_evidence](#)
- [analysis](#)
- [seq_region](#)

Stores the names of different genetic or radiation hybrid maps, for which there is marker map information.

Column	Type	Default value	Description	Index
map_id	INT(10)	-	Primary key, internal identifier.	primary key
map_name	VARCHAR(30)	-	Map name.	

See also:

List of species with populated data: [+ Show species](#)

[marker](#)

Stores data about the marker itself. A marker in Ensembl consists of a pair of primer sequences, an expected product size and a set of associated identifiers known as synonyms.

Column	Type	Default value	Description	Index
marker_id	INT(10)	-	Primary key, internal identifier.	primary key key: <i>marker_idx</i>
display_marker_synonym_id	INT(10)	NULL	Marker synonym.	key: <i>display_idx</i>
left_primer	VARCHAR(100)	-	Left primer sequence.	
right_primer	VARCHAR(100)	-	Right primer sequence.	
min_primer_dist	INT(10)	-	Minimum primer distance.	

max_primer_dist	INT(10)	-	Maximum primer distance.	
priority	INT	NULL	Priority.	key: <i>marker_idx</i>
type	ENUM: • est • microsatellite	NULL	Type, e.g. 'est', 'microsatellite'.	

See also: List of species with populated data: [+ Show species](#)

- [marker_synonym](#)
- [marker_map_location](#)

Used to describe positions of markers on the assembly. Markers are placed on the genome electronically using an analysis program.

Column	Type	Default value	Description	Index
marker_feature_id	INT(10)	-	Primary key, internal identifier.	primary key
marker_id	INT(10)	-	Foreign key references to the marker table.	
seq_region_id	INT(10)	-	Foreign key references to the seq_region table.	key: <i>seq_region_idx</i>
seq_region_start	INT(10)	-	Sequence start position.	key: <i>seq_region_idx</i>
seq_region_end	INT(10)	-	Sequence end position.	
analysis_id	SMALLINT	-	Foreign key references to the analysis table.	key: <i>analysis_idx</i>
map_weight	INT(10)	NULL	The number of times that this marker has been mapped to the genome, e.g. a marker with map weight 3 has been mapped to 3 locations in the genome.	

See also: List of species with populated data: [+ Show species](#)

- [marker](#)
- [marker_map_location](#)
- [marker_synonym](#)
- [seq_region](#)
- [analysis](#)

Stores map locations (genetic, radiation hybrid and in situ hybridization) for markers obtained from experimental evidence.

Column	Type	Default value	Description	Index
marker_id	INT(10)	-	Primary key, internal identifier.	primary key
map_id	INT(10)	-	Foreign key references to the map table.	primary key key: <i>map_idx</i>
chromosome_name	VARCHAR(15)	-	Chromosome name	key: <i>map_idx</i>
marker_synonym_id	INT(10)	-	Foreign key references to the marker_synonym table.	
position	VARCHAR(15)	-	Position of the map location.	key: <i>map_idx</i>
lod_score	DOUBLE	NULL	LOD score for map location.	

See also: List of species with populated data: [+ Show species](#)

- [marker](#)
- [marker_feature](#)
- [map](#)
- [marker_synonym](#)

Stores alternative names for markers, as well as their sources.

Column	Type	Default value	Description	Index
marker_synonym_id	INT(10)	-	Primary key, internal identifier.	primary key key: <i>marker_synonym_idx</i>
marker_id	INT(10)	-	Foreign key references to the <u>marker</u> table.	key: <i>marker_idx</i>
source	VARCHAR(20)	NULL	Marker source.	
name	VARCHAR(50)	NULL	Alternative name for marker.	key: <i>marker_synonym_idx</i>

See also: List of species with populated data: [+ Show species](#)
[marker](#)

Stores arbitrary attributes about the features in the misc_feature table.

Column	Type	Default value	Description	Index
misc_feature_id	INT(10)	'0'	Foreign key references to the <u>misc_feature</u> table.	key: <i>misc_feature_idx</i> unique key: <i>misc_attribx</i>
attrib_type_id	SMALLINT(5)	'0'	Foreign key references to the <u>attrib_type</u> table.	key: <i>type_val_idx</i> unique key: <i>misc_attribx</i>
value	TEXT	-	Attribute value.	key: <i>type_val_idx</i> key: <i>val_only_idx</i> unique key: <i>misc_attribx</i>

See also: List of species with populated data: [+ Show species](#)

- [misc_feature](#)
- [attrib_type](#)

Allows for storage of arbitrary features.

Column	Type	Default value	Description	Index
misc_feature_id	INT(10)	-	Primary key, internal identifier.	primary key
seq_region_id	INT(10)	'0'	Foreign key references to the <u>seq_region</u> table.	key: <i>seq_region_idx</i>
seq_region_start	INT(10)	'0'	Sequence start position.	key: <i>seq_region_idx</i>
seq_region_end	INT(10)	'0'	Sequence end position.	
seq_region_strand	TINYINT(4)	'0'	Sequence region strand: 1 - forward; -1 - reverse.	

See also: List of species with populated data: [+ Show species](#)

- [misc_attrib](#)
- [seq_region](#)

This table classifies features into distinct sets.

Column	Type	Default value	Description	Index
misc_feature_id	INT(10)	'0'	Primary key, internal identifier. Foreign key references to the <u>misc_feature</u> table.	primary key key: <i>reverse_idx</i>
misc_set_id	SMALLINT(5)	'0'	Primary key, internal identifier. Foreign key references to the <u>misc_feature</u> table.	primary key key: <i>reverse_idx</i>

See also: [List of species with populated data: !\[\]\(a3ea015cc5581cad732d1eb81613fe7b_img.jpg\) Show species](#)

- [misc_feature](#)
- [misc_set](#)

Defines "sets" that the features held in the misc_feature table can be grouped into.

Column	Type	Default value	Description	Index
misc_set_id	SMALLINT(5)	-	Primary key, internal identifier.	primary key
code	VARCHAR(25)	"	Set code, e.g. bac_map	unique key: <i>code_idx</i>
name	VARCHAR(255)	"	Code name, e.g. BAC map	
description	TEXT	-	Code description, e.g. Full list of mapped BAC clones	
max_length	INT	-	Longest feature, e.g. 500000	

See also: [List of species with populated data: !\[\]\(c3d993ca47bfe2a953c700506ce31fa0_img.jpg\) Show species](#)

[misc_feature_misc_set](#)

Stores exons that are predicted by ab initio gene finder programs. Unlike Ensembl exons they are not supported by any evidence.

Column	Type	Default value	Description	Index
prediction_exon_id	INT(10)	-	Primary key, internal identifier.	primary key
prediction_transcript_id	INT(10)	-	Foreign key references to the prediction_transcript table.	key: <i>transcript_idx</i>
exon_rank	SMALLINT	-	Exon rank	
seq_region_id	INT(10)	-	Foreign key references to the seq_region table.	key: <i>seq_region_idx</i>
seq_region_start	INT(10)	-	Sequence start position.	key: <i>seq_region_idx</i>
seq_region_end	INT(10)	-	Sequence end position.	
seq_region_strand	TINYINT	-	Sequence region strand: 1 - forward; -1 - reverse.	
start_phase	TINYINT	-	Exon start phase.	
score	DOUBLE	NULL	Prediction score.	
p_value	DOUBLE	NULL	Prediction p-value.	

See also: [List of species with populated data: !\[\]\(17413706fd4997a1a4bdf85c6864eee1_img.jpg\) Show species](#)

- [prediction_transcript](#)
- [seq_region](#)

Stores transcripts that are predicted by ab initio gene finder programs (e.g. genscan, SNAP). Unlike Ensembl transcripts they are not supported by any evidence.

Column	Type	Default value	Description	Index
prediction_transcript_id	INT(10)	-	Primary key, internal identifier.	primary key
seq_region_id	INT(10)	-	Foreign key references to the seq_region table.	key: <i>seq_region_idx</i>
seq_region_start	INT(10)	-	Sequence start position.	key: <i>seq_region_idx</i>
seq_region_end	INT(10)	-	Sequence end position.	
seq_region_strand	TINYINT	-	Sequence region strand: 1 - forward; -1 - reverse.	
analysis_id	SMALLINT	-	Foreign key references to the analysis table.	key: <i>analysis_idx</i>
display_label	VARCHAR(255)	NULL	Display label for the Ensembl web site.	

See also: [List of species with populated data: + Show species](#)

- [seq_region](#)
- [analysis](#)

Stores consensus sequences obtained from analysing repeat features.

Column	Type	Default value	Description	Index
repeat_consensus_id	INT(10)	-	Primary key, internal identifier.	primary key
repeat_name	VARCHAR(255)	-	Repeat name.	key: <i>name</i>
repeat_class	VARCHAR(100)	-	E.g. 'Satellite', 'tRNA', 'LTR'.	key: <i>class</i>
repeat_type	VARCHAR(40)	-	E.g. 'Satellite repeats', 'Tandem repeats', 'Low complexity regions'.	key: <i>type</i>
repeat_consensus	TEXT	NULL	Repeat consensus sequence.	key: <i>consensus</i>

List of species with populated data: [+ Show species](#)

Describes sequence repeat regions.

Column	Type	Default value	Description	Index
repeat_feature_id	INT(10)	-	Primary key, internal identifier.	primary key
seq_region_id	INT(10)	-	Foreign key references to the seq_region table.	key: <i>seq_region_idx</i>
seq_region_start	INT(10)	-	Sequence start position.	key: <i>seq_region_idx</i>
seq_region_end	INT(10)	-	Sequence end position.	
seq_region_strand	TINYINT(1)	'1'	Sequence region strand: 1 - forward; -1 - reverse.	
repeat_start	INT(10)	-	Repeat sequence start.	
repeat_end	INT(10)	-	Repeat sequence end	
repeat_consensus_id	INT(10)	-	Foreign key references to the repeat_consensus table.	key: <i>repeat_idx</i>
analysis_id	SMALLINT	-	Foreign key references to the analysis table.	key: <i>analysis_idx</i>
score	DOUBLE	NULL	Analysis score.	

See also: [List of species with populated data: + Show species](#)

- [seq_region](#)
- [repeat_consensus](#)
- [analysis](#)

Describes general genomic features that don't fit into any of the more specific feature tables.

Column	Type	Default value	Description	Index
simple_feature_id	INT(10)	-	Primary key, internal identifier.	primary key
seq_region_id	INT(10)	-	Foreign key references to the seq_region table.	key: <i>seq_region_idx</i>
seq_region_start	INT(10)	-	Sequence start position.	key: <i>seq_region_idx</i>
seq_region_end	INT(10)	-	Sequence end position.	
seq_region_strand	TINYINT(1)	-	Sequence region strand: 1 - forward; -1 - reverse.	
display_label	VARCHAR(255)	-	Display label for the Ensembl web site.	key: <i>hit_idx</i>
analysis_id	SMALLINT	-	Foreign key references to the analysis table.	key: <i>analysis_idx</i>

score	DOUBLE	NULL	Analysis score.
--------------	--------	------	-----------------

See also: List of species with populated data: [+ Show species](#)

- [seq_region](#)
- [analysis](#)

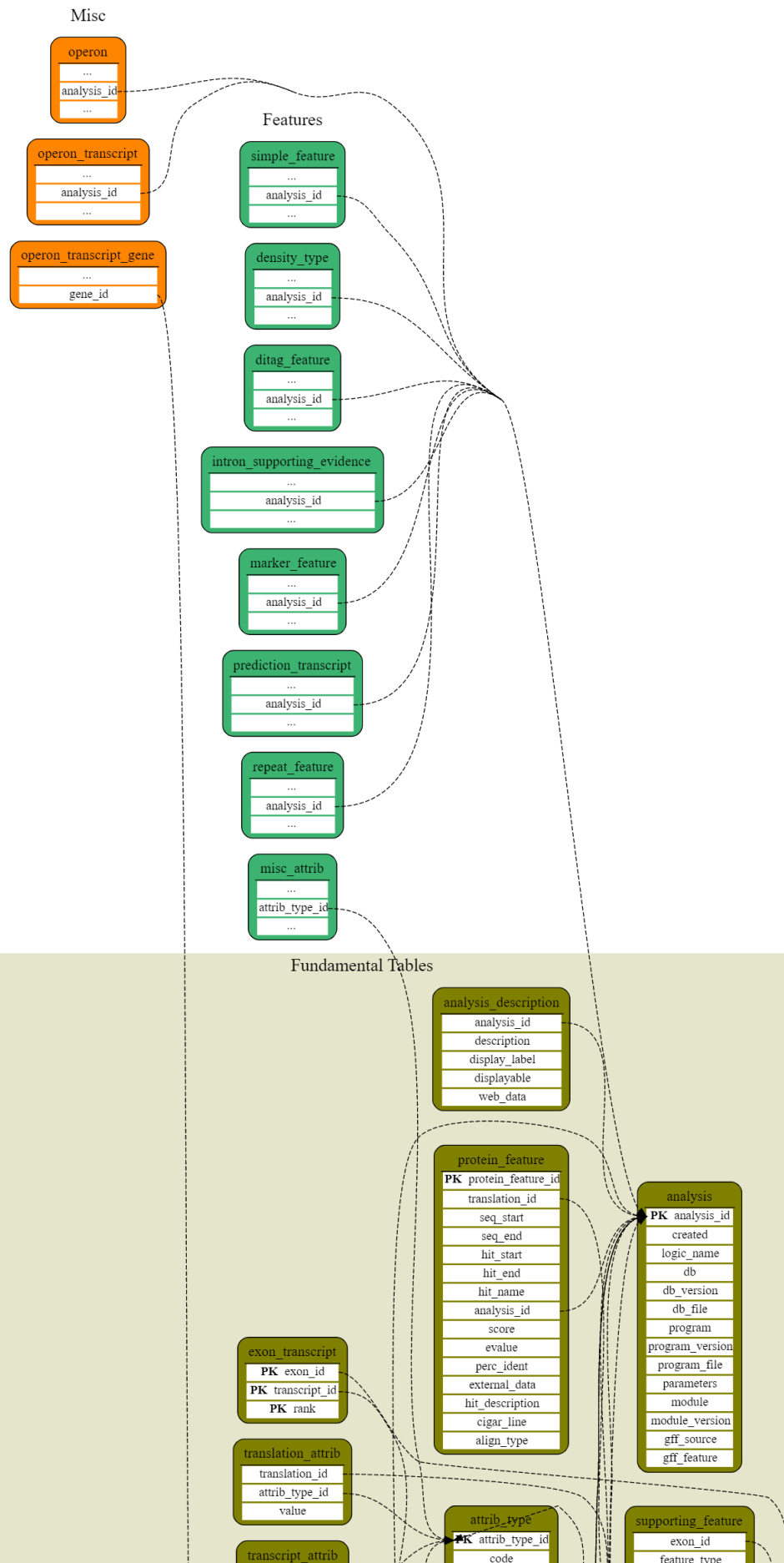
Links intronic evidence to a pair of exons used within a transcript and to resolve the m:m relationship between introns and transcripts

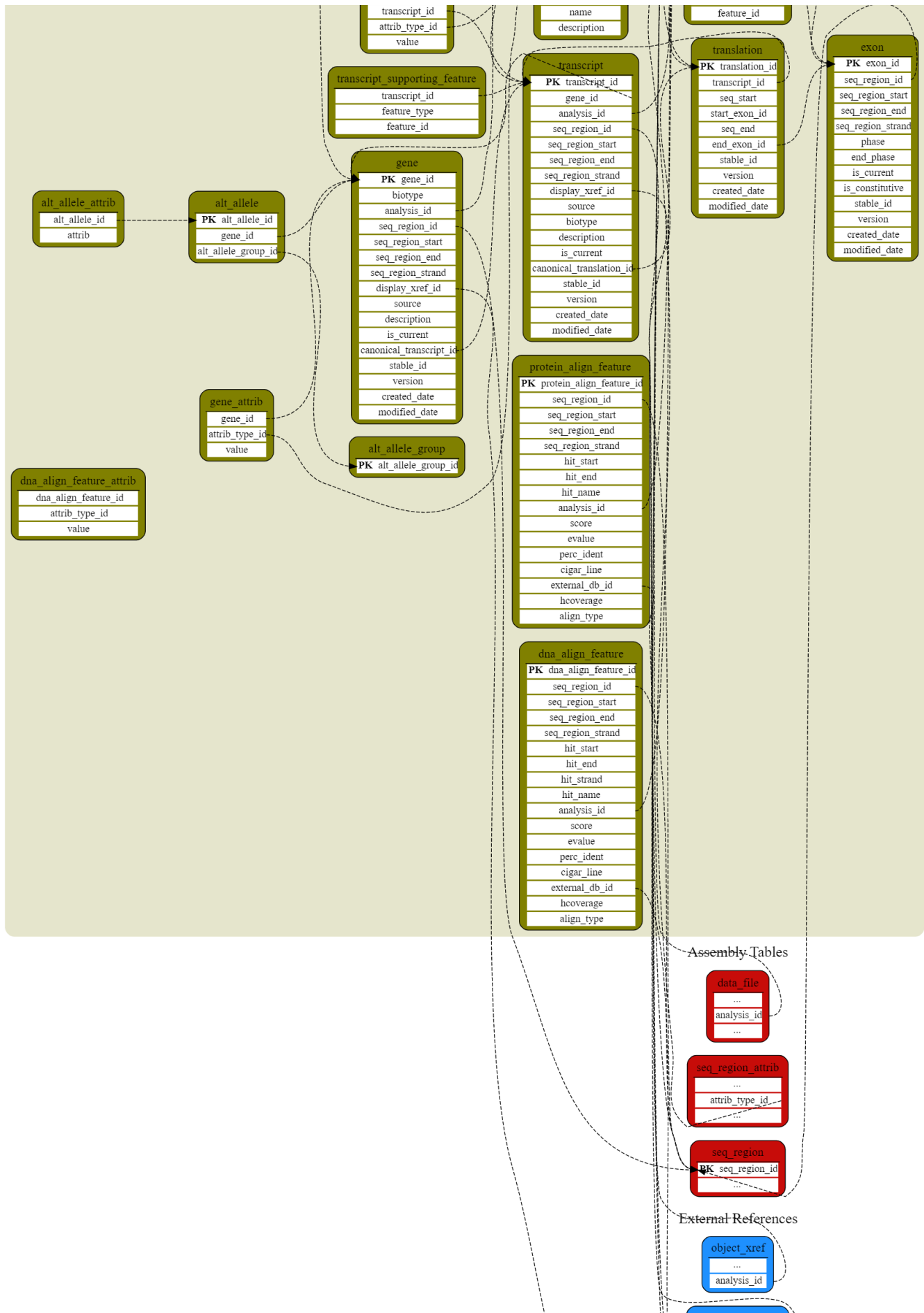
Column	Type	Default value	Description	Index
intron_supporting_evidence_id	INT(10)	-	Foreign key references to the intron_supporting_evidence table	primary key
transcript_id	INT(10)	-	Foreign key references to the transcript table.	primary key key: <i>transcript_idx</i>
previous_exon_id	INT(10)	-	Foreign key to exon indicating the left hand flanking exon of the intron (assume forward strand)	
next_exon_id	INT(10)	-	Foreign key to exon indicating the right hand flanking exon of the intron (assume forward strand)	

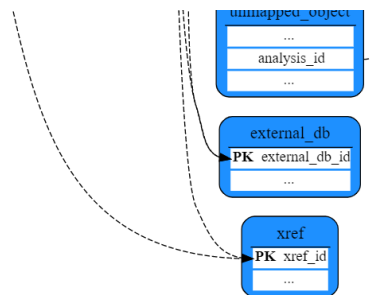
See also: List of species with populated data: [+ Show species](#)

- [intron_supporting_evidence](#)
- [transcript](#)
- [exon](#)

Fundamental Tables







Core schema diagram: Fundamental Tables tables

Stores information about genes on haplotypes that may be orthologous.

Column	Type	Default value	Description	Index
alt_allele_id	INT	NULL	Primary key, internal identifier.	primary key
gene_id	INT	-	Foreign key references to the <u>gene</u> table.	unique key: <i>gene_idx</i> key: <i>gene_id,alt_allele_group_id</i>
alt_allele_group_id	INT	-	A group ID to show which alleles are related	key: <i>gene_id,alt_allele_group_id</i>

See also: List of species with populated data: [+ Show species](#)
[gene](#)

Holds all the different attributes assigned to individual alleles.

Column	Type	Default value	Description	Index
alt_allele_id	INT	NULL	Primary key, internal identifier.	key: <i>aa_idx</i>
attrib	ENUM: <ul style="list-style-type: none"> IS_REPRESENTATIVE IS_MOST_COMMON_ALLELE IN_CORRECTED_ASSEMBLY HAS_CODING_POTENTIAL IN_ARTIFICIALLY_DUPLICATED_ASSEMBLY IN_SYNTENIC_REGION HAS_SAME_UNDERLYING_DNA_SEQUENCE IN_BROKEN_ASSEMBLY_REGION IS_VALID_ALTERNATE SAME_AS_REPRESENTATIVE SAME_AS_ANOTHER_ALLELE MANUALLY_ASSIGNED AUTOMATICALLY_ASSIGNED 	NULL	Enum of attributes assigned to alternative alleles	key: <i>aa_idx</i>

List of species with populated data: [+ Show species](#)

A minimal table used for tracking unique alt_allele_group_id's. MySQL does not allow multiple autoincrement fields. Further information about a group could be added here at a later date.

Column	Type	Default value	Description	Index
alt_allele_group_id	INT	NULL	Primary key and only column.	primary key

List of species with populated data: [+ Show species](#)

Usually describes a program and some database that together are used to create a feature on a piece of sequence. Each feature is marked with an analysis_id. The most important column is logic_name, which is used by the webteam to render a feature correctly on contigview (or even retrieve the right feature). Logic_name is also used in the pipeline to identify the analysis which has to run in a given status of the pipeline. The module column tells the pipeline which Perl module does the whole analysis, typically a RunnableDB module.

Column	Type	Default value	Description	Index
analysis_id	SMALLINT	-	Primary key, internal identifier.	primary key
created	datetime	NULL	Date to distinguish newer and older versions off the same analysis.	
logic_name	VARCHAR(128)	-	String to identify the analysis. Used mainly inside pipeline.	unique key: <i>logic_name_idx</i>
db	VARCHAR(120)	NULL	Database name.	
db_version	VARCHAR(40)	NULL	Database version.	
db_file	VARCHAR(120)	NULL	File system location of the database.	
program	VARCHAR(80)	NULL	The binary used to create a feature.	
program_version	VARCHAR(40)	NULL	The binary version.	
program_file	VARCHAR(80)	NULL	File system location of the binary.	
parameters	TEXT	NULL	A parameter string which is processed by the perl module.	
module	VARCHAR(80)	NULL	Perl module names (RunnableDBS usually) executing this analysis.	
module_version	VARCHAR(40)	NULL	Perl module version.	
gff_source	VARCHAR(40)	NULL	How to make a gff dump from features with this analysis.	
gff_feature	VARCHAR(40)	NULL	How to make a gff dump from features with this analysis.	

See also: [List of species with populated data: !\[\]\(35e4f762fc1cfea5610d92e2d225d5b4_img.jpg\) Show species](#)
[analysis_description](#)

Allows the storage of a textual description of the analysis, as well as a "display label", primarily for the Ensembl web site.

Column	Type	Default value	Description	Index
analysis_id	SMALLINT	-	Primary key, internal identifier. Foreign key references to the analysis table.	unique key: <i>analysis_idx</i>
description	TEXT	NULL	Textual description of the analysis.	
display_label	VARCHAR(255)	-	Display label for the Ensembl web site.	
displayable	TINYINT(1)	1	Flag indicating if the analysis description is to be displayed on the Ensembl web site.	
web_data	TEXT	NULL	Other data used by the Ensembl web site.	

See also: [List of species with populated data: !\[\]\(9dfdaff1d86ba3c1f8353b4d1b61b8c5_img.jpg\) Show species](#)
[analysis](#)

Provides codes, names and descriptions of attribute types.

Column	Type	Default value	Description	Index
attrib_type_id	SMALLINT(5)	-	Primary key, internal identifier.	primary key
code	VARCHAR(20)	"	Attribute code, e.g. 'GapExons'.	unique key: <i>code_idx</i>
name	VARCHAR(255)	"	Attribute name, e.g. 'gap exons'.	
description	TEXT	NULL	Attribute description, e.g. 'number of gap exons'.	

See also: [List of species with populated data: + Show species](#)

[seq_region_attrb](#)

Stores DNA sequence alignments generated from Blast (or Blast-like) comparisons.

Column	Type	Default value	Description	Index
dna_align_feature_id	INT(10)	-	Primary key, internal identifier.	primary key
seq_region_id	INT(10)	-	Foreign key references to the seq_region table.	key: seq_region_idx key: seq_region_idx_2
seq_region_start	INT(10)	-	Sequence start position.	key: seq_region_idx key: seq_region_idx_2
seq_region_end	INT(10)	-	Sequence end position.	
seq_region_strand	TINYINT(1)	-	Sequence region strand: 1 - forward; -1 - reverse.	
hit_start	INT	-	Alignment hit start position.	
hit_end	INT	-	Alignment hit end position.	
hit_strand	TINYINT(1)	-	Alignment hit strand: 1 - forward; -1 - reverse.	
hit_name	VARCHAR(40)	-	Alignment hit name.	key: hit_idx
analysis_id	SMALLINT	-	Foreign key references to the analysis table.	key: seq_region_idx key: analysis_idx
score	DOUBLE	NULL	Alignment score.	key: seq_region_idx
evalue	DOUBLE	NULL	Alignment e-value.	
perc_ident	FLOAT	NULL	Alignment percentage identity.	
cigar_line	TEXT	NULL	Used to encode gapped alignments.	
external_db_id	INT	NULL	Foreign key references to the external_db table.	key: external_db_idx
hcoverage	DOUBLE	NULL	Hit coverage.	
align_type	ENUM: <ul style="list-style-type: none">• ensembl• cigar• vulgar• mdtag	'ensembl'	Alignment string type used	

See also: [List of species with populated data: + Show species](#)

- [cigar_line](#)
- [seq_region](#)
- [analysis](#)
- [external_db](#)

Enables storage of attributes that relate to DNA sequence alignments.

Column	Type	Default value	Description	Index
dna_align_feature_id	INT(10)	-	Foreign key references to the dna_align_feature table.	unique key: dna_align_feature_attrbx key: dna_align_feature_idx
attrib_type_id	SMALLINT(5)	-	Foreign key references to the attrib_type table.	unique key: dna_align_feature_attrbx key: type_val_idx
value	TEXT	-	Attribute value.	unique key: dna_align_feature_attrbx key: type_val_idx key: val_only_idx

See also: [List of species with populated data: + Show species](#)

- [dna_align_feature](#)
- [attrib_type](#)

Stores data about exons. Associated with transcripts via exon_transcript. Allows access to contigs seq_regions. Note seq_region_start is always less than seq_region_end, i.e. when the exon is on the other strand the seq_region_start is specifying the 3prime end of the exon.

Column	Type	Default value	Description	Index
exon_id	INT(10)	-	Primary key, internal identifier.	primary key
seq_region_id	INT(10)	-	Foreign key references to the seq_region table.	key: seq_region_idx
seq_region_start	INT(10)	-	Sequence start position.	key: seq_region_idx
seq_region_end	INT(10)	-	Sequence end position.	
seq_region_strand	TINYINT(2)	-	Sequence region strand: 1 - forward; -1 - reverse.	
phase	TINYINT(2)	-	The place where the intron lands inside the codon - 0 between codons, 1 between the 1st and second base, 2 between the second and 3rd base. Exons therefore have a start phase and an end phase, but introns have just one phase.	
end_phase	TINYINT(2)	-	Usually, end_phase = (phase + exon_length)%3 but end_phase could be -1 if the exon is half-coding and its 3 prime end is UTR.	
is_current	TINYINT(1)	1	1 - exon is current. Always set to 1 in ensembl dbs, but needed for otterlace dbs	
is_constitutive	TINYINT(1)	0	1 - exon is constitutive.	
stable_id	VARCHAR(128)	NULL	Release-independent stable identifier.	key: stable_id_idx
version	SMALLINT	NULL	Stable identifier version number.	key: stable_id_idx
created_date	DATETIME	NULL	Date created.	
modified_date	DATETIME	NULL	Date modified.	

See also: [List of species with populated data: + Show species](#)

- [exon_transcript](#)
- [seq_region](#)

Relationship table linking exons with transcripts. The rank column indicates the 5' to 3' position of the exon within the transcript, i.e. a rank of 1 means the exon is the 5' most within this transcript.

Column	Type	Default value	Description	Index
exon_id	INT(10)	-	Composite key. Foreign key references to the exon table.	primary key key: exon
transcript_id	INT(10)	-	Composite key. Foreign key references to the transcript table.	primary key key: transcript
rank	INT(10)	-	Composite key.	primary key

See also: [List of species with populated data: + Show species](#)

- [exon](#)
- [transcript](#)

Allows transcripts to be related to genes.

Column	Type	Default value	Description	Index
gene_id	INT(10)	-	Primary key, internal identifier.	primary key
biotype	VARCHAR(40)	-	Biotype, e.g. protein_coding.	
analysis_id	SMALLINT	-	Foreign key references to the analysis table.	key: <i>analysis_idx</i>
seq_region_id	INT(10)	-	Foreign key references to the seq_region table.	key: <i>seq_region_idx</i>
seq_region_start	INT(10)	-	Sequence start position.	key: <i>seq_region_idx</i>
seq_region_end	INT(10)	-	Sequence end position.	
seq_region_strand	TINYINT(2)	-	Sequence region strand: 1 - forward; -1 - reverse.	
display_xref_id	INT(10)	NULL	External reference for Ensembl web site. Foreign key references to the xref table.	key: <i>xref_id_index</i>
source	VARCHAR(40)	-	e.g ensembl, havana etc.	
description	TEXT	NULL	Gene description	
is_current	TINYINT(1)	1	1 - gene is current. Always set to 1 in ensembl dbs, but needed for otterlace dbs	
canonical_transcript_id	INT(10)	-	Foreign key references to the transcript table.	key: <i>canonical_transcript_id_idx</i>
stable_id	VARCHAR(128)	NULL	Release-independent stable identifier.	key: <i>stable_id_idx</i>
version	SMALLINT	NULL	Stable identifier version number.	key: <i>stable_id_idx</i>
created_date	DATETIME	NULL	Date created.	
modified_date	DATETIME	NULL	Date modified.	

See also: [List of species with populated data: !\[\]\(529949c2c3dadbaa4e538e8c643454bc_img.jpg\) \[Show species\]\(#\)](#)

- [transcript](#)
- [analysis](#)
- [seq_region](#)
- [xref](#)

Enables storage of attributes that relate to genes.

Column	Type	Default value	Description	Index
gene_id	INT(10)	'0'	Foreign key references to the gene table.	key: <i>gene_idx</i> unique key: <i>gene_attribx</i>
attrib_type_id	SMALLINT(5)	'0'	Foreign key references to the attrib_type table.	key: <i>type_val_idx</i> unique key: <i>gene_attribx</i>
value	TEXT	-	Attribute value.	key: <i>type_val_idx</i> key: <i>val_only_idx</i> unique key: <i>gene_attribx</i>

See also: [List of species with populated data: !\[\]\(339a16584d5da0f0a3ca4e9ec17bf6a1_img.jpg\) \[Show species\]\(#\)](#)

- [gene](#)
- [attrib_type](#)

Stores translation alignments generated from Blast (or Blast-like) comparisons.

Column	Type	Default value	Description	Index
protein_align_feature_id	INT(10)	-	Primary key, internal identifier.	primary key
seq_region_id	INT(10)	-	Foreign key references to the seq_region table.	key: <i>seq_region_idx</i> key: <i>seq_region_idx_2</i>

seq_region_start	INT(10)	-	Sequence start position.	key: <i>seq_region_idx</i> key: <i>seq_region_idx_2</i>
seq_region_end	INT(10)	-	Sequence end position.	
seq_region_strand	TINYINT(1)	'1'	Sequence region strand: 1 - forward; -1 - reverse.	
hit_start	INT(10)	-	Alignment hit start position.	
hit_end	INT(10)	-	Alignment hit end position.	
hit_name	VARCHAR(40)	-	Alignment hit name.	key: <i>hit_idx</i>
analysis_id	SMALLINT	-	Foreign key references to the <u>analysis</u> table.	key: <i>seq_region_idx</i> key: <i>analysis_idx</i>
score	DOUBLE	NULL	Alignment score.	key: <i>seq_region_idx</i>
evalue	DOUBLE	NULL	Alignment e-value.	
perc_ident	FLOAT	NULL	Alignment percentage identity.	
cigar_line	TEXT	NULL	Used to encode gapped alignments.	
external_db_id	INT	NULL	Foreign key references to the <u>external_db</u> table.	key: <i>external_db_idx</i>
hcoverage	DOUBLE	NULL	Alignment hit coverage.	
align_type	ENUM: • ensembl • cigar • vulgar • mdtag	'ensembl'	Alignment string type used	

See also: [List of species with populated data: !\[\]\(1d3a1175dd4902218e694b9c098adb83_img.jpg\) Show species](#)

- [cigar_line](#)
- [seq_region](#)
- [analysis](#)
- [external_db](#)

Describes features on the translations (as opposed to the DNA sequence itself), i.e. parts of the peptide. In peptide co-ordinates rather than contig co-ordinates.

Column	Type	Default value	Description	Index
protein_feature_id	INT(10)	-	Primary key, internal identifier.	primary key
translation_id	INT(10)	-	Foreign key references to the <u>translation</u> table.	unique key: <i>aln_idx</i> key: <i>translation_idx</i>
seq_start	INT(10)	-	Sequence start position.	unique key: <i>aln_idx</i>
seq_end	INT(10)	-	Sequence end position.	unique key: <i>aln_idx</i>
hit_start	INT(10)	-	Alignment hit start position.	unique key: <i>aln_idx</i>
hit_end	INT(10)	-	Alignment hit end position.	unique key: <i>aln_idx</i>
hit_name	VARCHAR(40)	-	Alignment hit name.	unique key: <i>aln_idx</i> key: <i>hitname_idx</i>
analysis_id	SMALLINT	-	Foreign key references to the <u>analysis</u> table.	unique key: <i>aln_idx</i> key: <i>analysis_idx</i>
score	DOUBLE	NULL	Alignment score.	

evalue	DOUBLE	NULL	Alignment E-value.
perc_ident	FLOAT	NULL	Alignment percentage identity.
external_data	TEXT	NULL	External data for protein feature.
hit_description	TEXT	NULL	Optional description of the hit. This can be a human readable name
cigar_line	TEXT	NULL	Used to encode gapped alignments.
align_type	ENUM: • ensembl • cigar • cigarplus • vulgar • mdtag	NULL	Alignment string type used

See also: List of species with populated data: [+ Show species](#)

- [analysis](#)
- [translation](#)

Describes the exon prediction process by linking exons to DNA or protein alignment features. As in several other tables, the feature_id column is a foreign key; the feature_type column specifies which table feature_id refers to.

Column	Type	Default value	Description	Index
exon_id	INT(10)	'0'	Foreign key references to the exon table.	unique key: <i>all_idx</i>
feature_type	ENUM: • dna_align_feature • protein_align_feature	NULL	Feature type: 'dna_align_feature' or 'protein_align_feature'	unique key: <i>all_idx</i> key: <i>feature_idx</i>
feature_id	INT(10)	'0'	Foreign key references to the dna_align_feature or protein_align_feature table depending on the feature type.	unique key: <i>all_idx</i> key: <i>feature_idx</i>

See also: List of species with populated data: [+ Show species](#)

- [exon](#)
- [dna_align_feature](#)
- [protein_align_feature](#)

Stores information about transcripts. Has seq_region_start, seq_region_end and seq_region_strand for faster retrieval and to allow storage independently of genes and exons. Note that a transcript is usually associated with a translation, but may not be, e.g. in the case of pseudogenes and RNA genes (those that code for RNA molecules).

Column	Type	Default value	Description	Index
transcript_id	INT(10)	-	Primary key, internal identifier.	primary key
gene_id	INT(10)	NULL	Foreign key references to the gene table.	key: <i>gene_index</i>
analysis_id	SMALLINT	-	Foreign key references to the analysis table.	key: <i>analysis_idx</i>
seq_region_id	INT(10)	-	Foreign key references to the seq_region table.	key: <i>seq_region_idx</i>
seq_region_start	INT(10)	-	Sequence start position.	key: <i>seq_region_idx</i>
seq_region_end	INT(10)	-	Sequence end position.	
seq_region_strand	TINYINT(2)	-	Sequence region strand: 1 - forward; -1 - reverse.	
display_xref_id	INT(10)	NULL	External reference for EnSEMBL web site. Foreign key references to the xref table.	key: <i>xref_id_index</i>

source	VARCHAR(40)	'ensembl'	e.g ensembl, havana etc.	
biotype	VARCHAR(40)	-	Biotype, e.g. protein_coding.	
description	TEXT	NULL	Transcript description.	
is_current	TINYINT(1)	1	Indicates a current transcript. Always set to 1 in ensembl dbs, but needed for otterlace dbs	
canonical_translation_id	INT(10)	NULL	Foreign key references to the translation table.	unique key: <i>canonical_translation_idx</i>
stable_id	VARCHAR(128)	NULL	Release-independent stable identifier.	key: <i>stable_id_idx</i>
version	SMALLINT	NULL	Stable identifier version number.	key: <i>stable_id_idx</i>
created_date	DATETIME	NULL	Date created.	
modified_date	DATETIME	NULL	Date modified.	

See also: List of species with populated data: [+ Show species](#)

- [gene](#)
- [analysis](#)
- [seq_region](#)
- [xref](#)
- [translation](#)

Enables storage of attributes that relate to transcripts.

Column	Type	Default value	Description	Index
transcript_id	INT(10)	'0'	Foreign key references to the transcript table.	key: <i>transcript_idx</i> unique key: <i>transcript_attridx</i>
attrib_type_id	SMALLINT(5)	'0'	Foreign key references to the attrib_type table.	key: <i>type_val_idx</i> unique key: <i>transcript_attridx</i>
value	TEXT	-	Attribute value.	key: <i>type_val_idx</i> key: <i>val_only_idx</i> unique key: <i>transcript_attridx</i>

See also: List of species with populated data: [+ Show species](#)

- [transcript](#)
- [attrib_type](#)

Describes the exon prediction process by linking transcripts to DNA or protein alignment features. As in several other tables, the feature_id column is a foreign key; the feature_type column specifies which table feature_id refers to.

Column	Type	Default value	Description	Index
transcript_id	INT(10)	'0'	Foreign key references to the transcript table.	unique key: <i>all_idx</i>
feature_type	ENUM: • dna_align_feature • protein_align_feature	NULL	Feature type: 'dna_align_feature' or 'protein_align_feature'	unique key: <i>all_idx</i> key: <i>feature_idx</i>
feature_id	INT(10)	'0'	Foreign key references to the dna_align_feature or protein_align_feature table depending on the feature type.	unique key: <i>all_idx</i> key: <i>feature_idx</i>

See also: List of species with populated data: [+ Show species](#)

- [transcript](#)
- [dna_align_feature](#)
- [protein_align_feature](#)

Describes which parts of which exons are used in translation. The seq_start and seq_end columns are 1-based offsets into the relative coordinate system of start_exon_id and end_exon_id. i.e, if the translation starts at the first base of the exon, seq_start would be 1. Transcripts are related to translations by the transcript_id key in this table.

Column	Type	Default value	Description	Index
translation_id	INT(10)	-	Primary key, internal identifier.	primary key
transcript_id	INT(10)	-	Foreign key references to the transcript table.	key: <i>transcript_idx</i>
seq_start	INT(10)	-	1-based offset into the relative coordinate system of start_exon_id.	
start_exon_id	INT(10)	-	Foreign key references to the exon table.	
seq_end	INT(10)	-	1-based offset into the relative coordinate system of end_exon_id.	
end_exon_id	INT(10)	-	Foreign key references to the exon table.	
stable_id	VARCHAR(128)	NULL	Release-independent stable identifier.	key: <i>stable_id_idx</i>
version	SMALLINT	NULL	Stable identifier version number.	key: <i>stable_id_idx</i>
created_date	DATETIME	NULL	Date created.	
modified_date	DATETIME	NULL	Date modified.	

See also: [List of species with populated data:](#) [+ Show species](#)

- [transcript](#)
- [exon](#)

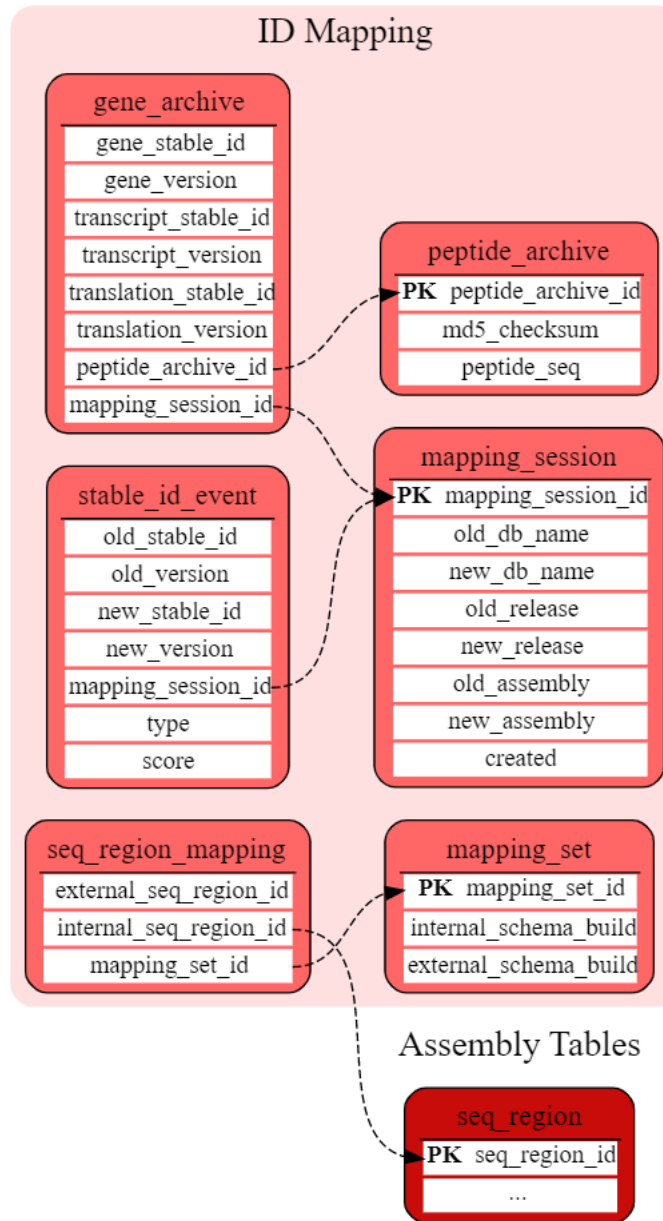
Enables storage of attributes that relate to translations.

Column	Type	Default value	Description	Index
translation_id	INT(10)	'0'	Foreign key references to the transcript table.	key: <i>translation_idx</i> unique key: <i>translation_attribx</i>
attrib_type_id	SMALLINT(5)	'0'	Foreign key references to the attrib_type table.	key: <i>type_val_idx</i> unique key: <i>translation_attribx</i>
value	TEXT	-	Attribute value.	key: <i>type_val_idx</i> key: <i>val_only_idx</i> unique key: <i>translation_attribx</i>

See also: [List of species with populated data:](#) [+ Show species](#)

- [translation](#)
- [transcript](#)
- [attrib_type](#)

ID Mapping



Core schema diagram: ID Mapping tables

Contains a snapshot of the stable IDs associated with genes deleted or changed between releases. Includes gene, transcript and translation stable IDs.

Column	Type	Default value	Description	Index
gene_stable_id	VARCHAR(128)	-	Stable ID of retired gene.	key: <i>gene_idx</i>
gene_version	SMALLINT	1	Last live gene stable ID version.	key: <i>gene_idx</i>
transcript_stable_id	VARCHAR(128)	-	Stable ID of associated transcript.	key: <i>transcript_idx</i>
transcript_version	SMALLINT	1	Last live transcript stable ID version.	key: <i>transcript_idx</i>
translation_stable_id	VARCHAR(128)	NULL	Stable ID of associated translation.	key: <i>translation_idx</i>
translation_version	SMALLINT	1	Last live translation stable ID.	key: <i>translation_idx</i>
peptide_archive_id	INT(10)	NULL	Foreign key references to the <u>peptide</u> archive table.	key: <i>peptide_archive_id_idx</i>

mapping_session_id	INT(10)	-	Foreign key references to the <u>mapping_session</u> table.
---------------------------	---------	---	---

See also: [List of species with populated data: !\[\]\(d263118e0bfd47dc6bc704167d936b83_img.jpg\) Show species](#)

- [gene](#)
- [peptide](#)
- [mapping_session](#)

Stores details of ID mapping sessions - a mapping session represents the session when stable IDs were mapped from one database to another. Details of the "old" and "new" databases are stored.

Column	Type	Default value	Description	Index
mapping_session_id	INT(10)	-	Primary key, internal identifier.	primary key
old_db_name	VARCHAR(80)	"	Old Ensembl database name.	
new_db_name	VARCHAR(80)	"	New Ensembl database name.	
old_release	VARCHAR(5)	"	Old Ensembl database release.	
new_release	VARCHAR(5)	"	New Ensembl database release.	
old_assembly	VARCHAR(80)	"	Old assembly.	
new_assembly	VARCHAR(80)	"	New assembly.	
created	DATETIME	-	Date created.	

See also: [List of species with populated data: !\[\]\(96cc62f861fdd6e50510c0224a756dff_img.jpg\) Show species](#)

- [stable_id_event](#)
- [stable_id](#)

Table structure for seq_region mapping between releases.

Column	Type	Default value	Description	Index
mapping_set_id	INT(10)	-	Primary key, internal identifier.	primary key
internal_schema_build	VARCHAR(20)	-	Schema version of the current database (eg 72_37)	unique key: <i>mapping_idx</i>
external_schema_build	VARCHAR(20)	-	Schema version of the database the comparison was run against (eg 71_37)	unique key: <i>mapping_idx</i>

List of species with populated data: [+ Show species](#)

Contains the peptides for deleted or changed translations.

Column	Type	Default value	Description	Index
peptide_archive_id	INT(10)	-	Primary key, internal identifier.	primary key
md5_checksum	VARCHAR(32)	NULL	MD5 checksum hexadecimal digest of the peptide sequence.	key: <i>checksum</i>
peptide_seq	MEDIUMTEXT	-	Peptide sequence of retired translation.	

List of species with populated data: [+ Show species](#)

Describes how the core seq_region_id have changed from release to release.

Column	Type	Default value	Description	Index
external_seq_region_id	INT(10)	-	Foreign key references to the <u>seq_region</u> table.	

internal_seq_region_id	INT(10)	-	Foreign key references to the seq_region table.
mapping_set_id	INT(10)	-	Foreign key references to the mapping_set table. key: <i>mapping_set_idx</i>

See also: [List of species with populated data:](#) [+ Show species](#)

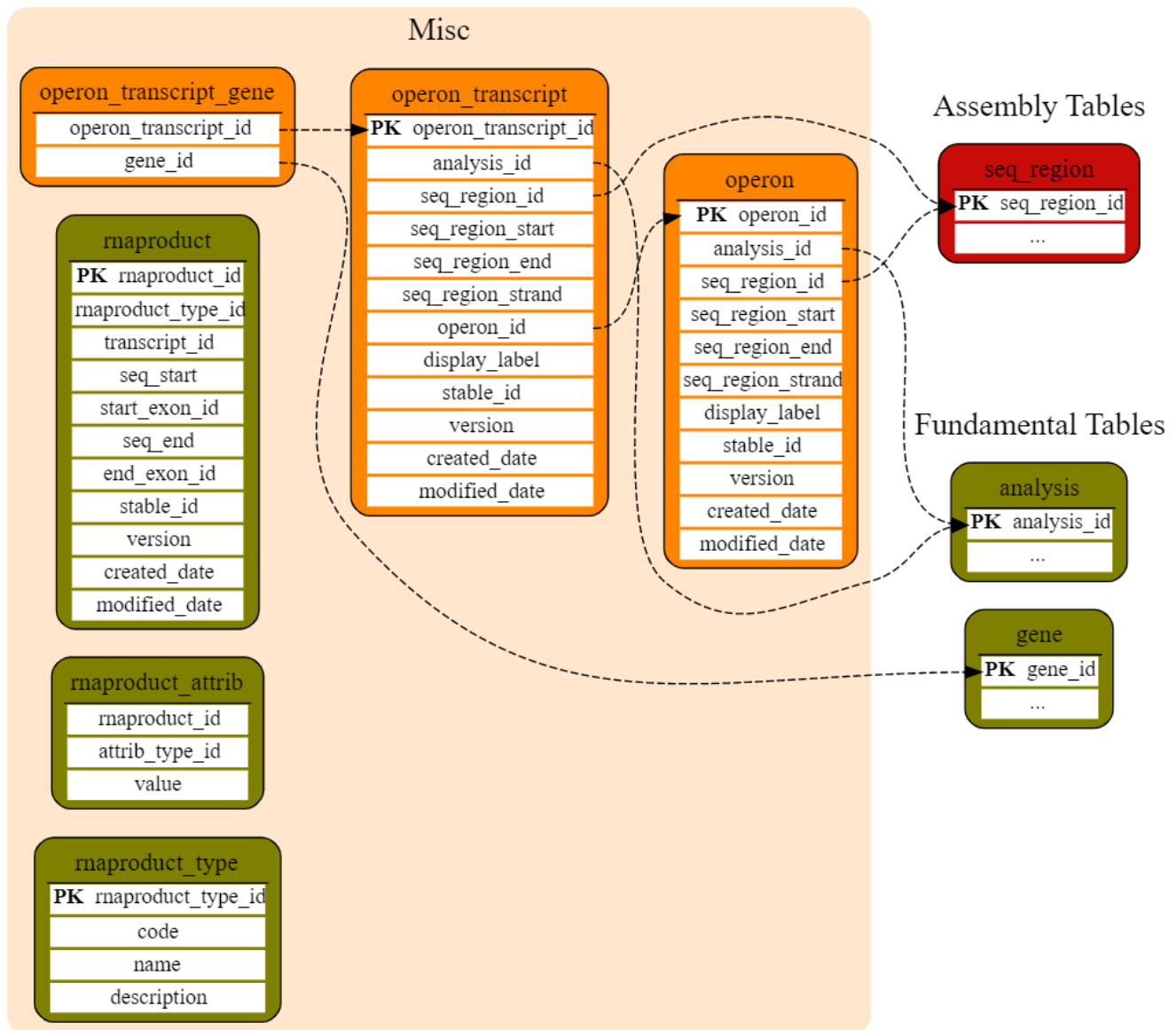
- [seq_region](#)
- [mapping_set](#)

Represents what happened to all gene, transcript and translation stable IDs during a mapping session. This includes which IDs were deleted, created and related to each other. Each event is represented by one or more rows in the table.

Column	Type	Default value	Description	Index
old_stable_id	VARCHAR(128)	NULL	Gene/transcript/translation stable id for the previous release.	unique key: <i>uni_idx</i> key: <i>old_idx</i>
old_version	SMALLINT	NULL	Stable id version.	
new_stable_id	VARCHAR(128)	NULL	Gene/transcript/translation stable id for the current release.	unique key: <i>uni_idx</i> key: <i>new_idx</i>
new_version	SMALLINT	NULL	Stable id version.	
mapping_session_id	INT(10)	'0'	Foreign key references to the mapping_session table.	unique key: <i>uni_idx</i>
type	ENUM: • gene • transcript • translation • mapproduct	-	ENUM('gene', 'transcript', 'translation', 'mapproduct') NOT NULL,	unique key: <i>uni_idx</i>
score	FLOAT	0	Combined mapping score.	

See also: [List of species with populated data:](#) [+ Show species](#)
[mapping_session](#)

Misc



Core schema diagram: Misc tables

allows one or more polycistronic transcripts to be grouped together

Column	Type	Default value	Description	Index
operon_id	INT(10)	-	Primary key, internal identifier.	primary key
analysis_id	SMALLINT	-	Foreign key references to the <u>analysis</u> table.	
seq_region_id	INT(10)	-	Foreign key references to the <u>seq_region</u> table.	key: <i>seq_region_idx</i>
seq_region_start	INT(10)	-	Sequence start position.	key: <i>seq_region_idx</i>
seq_region_end	INT(10)	-	Sequence end position.	
seq_region_strand	TINYINT(2)	-	Sequence region strand: 1 - forward; -1 - reverse.	
display_label	VARCHAR(255)	NULL	Short name for operon	key: <i>name_idx</i>
stable_id	VARCHAR(128)	NULL	Release-independent stable identifier.	key: <i>stable_id_idx</i>
version	SMALLINT	NULL	Stable identifier version number.	key: <i>stable_id_idx</i>
created_date	DATETIME	NULL	Date created.	
modified_date	DATETIME	NULL	Date modified.	

See also:

- [operon_transcript](#)
- [operon_stable_id](#)
- [analysis](#)
- [seq_region](#)

represents polycistronic transcripts which belong to operons and encode more than one gene

Column	Type	Default value	Description	Index
operon_transcript_id	INT(10)	-	Primary key, internal identifier.	primary key
analysis_id	SMALLINT	-	Foreign key references to the analysis table.	
seq_region_id	INT(10)	-	Foreign key references to the seq_region table.	key: <i>seq_region_idx</i>
seq_region_start	INT(10)	-	Sequence start position.	key: <i>seq_region_idx</i>
seq_region_end	INT(10)	-	Sequence end position.	
seq_region_strand	TINYINT(2)	-	Sequence region strand: 1 - forward; -1 - reverse.	
operon_id	INT(10)	-	Foreign key references to the operon table.	key: <i>operon_idx</i>
display_label	VARCHAR(255)	NULL	Short name for operon transcript	
stable_id	VARCHAR(128)	NULL	Release-independent stable identifier.	key: <i>stable_id_idx</i>
version	SMALLINT	NULL	Stable identifier version number.	key: <i>stable_id_idx</i>
created_date	DATETIME	NULL	Date created.	
modified_date	DATETIME	NULL	Date modified.	

See also:

- [operon](#)
- [operon_transcript_stable_id](#)
- [operon_transcript_gene](#)
- [analysis](#)
- [seq_region](#)

allows association of genes with polycistronic transcripts

Column	Type	Default value	Description	Index
operon_transcript_id	INT(10)	NULL	Foreign key references to the operon_transcript table.	key: <i>operon_transcript_gene_idx</i>
gene_id	INT(10)	NULL	Foreign key references to the gene table.	key: <i>operon_transcript_gene_idx</i>

See also:

- [operon_transcript](#)
- [gene](#)

Describes which parts of which precursor transcript are used in rnaproduct. The seq_start and seq_end columns are 1-based offsets into the relative coordinate system of transcript_id. i.e, if the rnaproduct starts at the first base of the transcript, seq_start would be 1. Transcripts are related to rnaproducts by the transcript_id key in this table.

Column	Type	Default value	Description	Index
rnaproduct_id	INT(10)	-	Primary key, internal identifier.	primary key
rnaproduct_type_id	SMALLINT(5)	-	Foreign key references to the rnaproduct_type table.	
transcript_id	INT(10)	-	Foreign key references to the transcript table.	key: <i>transcript_idx</i>

seq_start	INT(10)	-	1-based offset into the relative coordinate system of transcript_id.	
start_exon_id	INT(10)	NULL	Foreign key references to the exon table.	
seq_end	INT(10)	-	1-based offset into the relative coordinate system of transcript_id.	
end_exon_id	INT(10)	NULL	Foreign key references to the exon table.	
stable_id	VARCHAR(128)	NULL	Release-independent stable identifier.	key: <i>stable_id_idx</i>
version	SMALLINT	NULL	Stable identifier version number.	key: <i>stable_id_idx</i>
created_date	DATETIME	NULL	Date created.	
modified_date	DATETIME	NULL	Date modified.	

See also:

- [rnaproduct_type](#)
- [transcript](#)
- [exon](#)

Enables storage of attributes that relate to rnaproducts.

Column	Type	Default value	Description	Index
rnaproduct_id	INT(10)	'0'	Foreign key references to the transcript table.	key: <i>rnaproduct_idx</i> unique key: <i>rnaproduct_attribx</i>
attrib_type_id	SMALLINT(5)	'0'	Foreign key references to the attrib_type table.	key: <i>type_val_idx</i> unique key: <i>rnaproduct_attribx</i>
value	TEXT	-	Attribute value.	key: <i>type_val_idx</i> key: <i>val_only_idx</i> unique key: <i>rnaproduct_attribx</i>

See also:

- [rnaproduct](#)
- [transcript](#)
- [attrib_type](#)

Provides codes, names and descriptions of rnaproduct types.

Column	Type	Default value	Description	Index
rnaproduct_type_id	SMALLINT(5)	-	Primary key, internal identifier.	primary key
code	VARCHAR(20)	"	Attribute code, e.g. 'miRNA'.	unique key: <i>code_idx</i>
name	VARCHAR(255)	"	Attribute name, e.g. 'microRNA'.	
description	TEXT	NULL	Attribute description, e.g. 'mature microRNA'.	

See also:

[seq_region_rnaproduct](#)