Compara database is coupled to Ensembl core databases

Compara stores relationships between the genomes by loading references or ‘handles’ to external data.

Since there is minimal primary data inside Compara, to gain full access to the data these external links must be re-established

Example: compara_70 must be linked with the Ensembl core_70 databases

Proper REGISTRY configuration is needed
Alignments are stored in the genomic_align and genomic_align_block tables

A small example:

```
gorilla_gorilla/MT/935-953   gacat-ttaactaaacccc
macaca_mulatta/MT/1469-1488  aacatcttaactaaacgccc
pan_troglodytes/MT/934-953   gatac-ttaacttaaaaccccc
pongo_pygmaeus/MT/940-958    actac-ctaacttaaaac-ccc
homo_sapiens/MT/1516-1534   gacat-ttaactaaacccc
*   ***** **   ***
```

```
<table>
<thead>
<tr>
<th>Sequence</th>
<th>ID</th>
</tr>
</thead>
<tbody>
<tr>
<td>GACATTTAACTAAAACCCCA</td>
<td>5MD11MD3M</td>
</tr>
<tr>
<td>AACATCTTTAACTAAACGCCC</td>
<td>17MD3M</td>
</tr>
<tr>
<td>GATACCTTAACTTTAAACCCCA</td>
<td>5MD15M</td>
</tr>
<tr>
<td>ACTACCTTAACTTTAAACCCCA</td>
<td>5MD11MD3M</td>
</tr>
<tr>
<td>GACATTTAACTTTAAACCCCA</td>
<td>5MD11MD3M</td>
</tr>
</tbody>
</table>
```

5 genomic_align entries
1 genomic_align_block

Sequences from core
Exercise genome_align_block

A GenomicAlignBlock represents an alignment between two or more regions of genomic DNA. Within these blocks every region of genomic DNA is represented by a GenomicAlign object.

a) Print the LASTZ-NET alignments for pig chromosome 15 with cow (using pig coordinates 105734307 and 105739335).

Hint: use the MethodLinkSpeciesSet adaptor fetch_by_method_link_type_registry_aliases() method (eg. $mlss_adaptor->fetch_by_method_link_type_registry_aliases("LASTZ_NET", ['pig', "cow"])).

You will need a *core* slice for pig chromosome 15.

Get a GenomicAlignBlock adaptor and use the fetch_all_by_MethodLinkSpeciesSet_Slice() method to retrieve the genomic_align_block object(s).

Use the GenomicAlignBlock restrict_between_reference_positions() method to "restrict" the genomic_align_block between base pairs 105734307 and 105739335.

Finally print the alignment using the GenomicAlignBlock get_SimpleAlign() method (use AlignIO to format the output alignment).
genome_align_block

Chr15
genome_align_block

Chr15

1000

2000
genome_align_block

Chr15

1000

2000
genome_align_block
genome_align_block
Sample code genome_align_blocks

# get a list-ref of (data objects) genomic_align_blocks
my $all_genomic_align_blocks = $genomic_align_block_adaptor-
> fetch_all_by_MethodLinkSpeciesSet_Slice(
    $methodLinkSpeciesSet, $human_slice);

foreach my $genomic_align_block (@{$all_genomic_align_blocks}) {
    #print Dumper $genomic_align_block;
    my $restricted_genomic_align_block = $genomic_align_block-
> restrict_between_alignment_positions($ref_start, $ref_end);

    # get some information from the data object
    print $alignIO $restricted_genomic_align_block-
> get_SimpleAlign;
}
A GenomicAlignBlock represents an alignment between two or more regions of genomic DNA. Within these blocks every region of genomic DNA is represented by a GenomicAlign object.

a) Print the LASTZ-NET alignments for pig chromosome 15 with cow (using pig coordinates 105734307 and 105739335).

Hint: use the MethodLinkSpeciesSet adaptor fetch_by_method_link_type_registry_aliases() method (eg. $mlss_adaptor->fetch_by_method_link_type_registry_aliases("LASTZ_NET", ["pig", "cow"])).

You will need a *core* slice for pig chromosome 15.

Get a GenomicAlignBlock adaptor and use the fetch_all_by_MethodLinkSpeciesSet_Slice() method to retrieve the genomic_align_block object(s).

Use the GenomicAlignBlock restrict_between_reference_positions() method to "restrict" the genomic_align_block between base pairs 105734307 and 105739335.

Finally print the alignment using the GenomicAlignBlock get_SimpleAlign() method (use AlignIO to format the output alignment).
Exercise genome_align_block

b) Change the above example so that it prints the 13-way eutherian mammal (EPO) multiple alignments.

Hint: use the MethodLinkSpeciesSet adaptor fetch_by_method_link_type_species_set_name() method (the method_link_type is "EPO" and the species_set_name is "mammals").