

Bio::DB::GenBank

- ▶ Zugriff auf Sequenzen der NCBI GenBank
<http://www.ncbi.nlm.nih.gov/Genbank/index.html>
- ▶ Hier beispielhaft für Zugriffe auf Sequenzdatenbanken

Konstruktor:

```
$gb = Bio::DB::GenBank->new();
```

Erstellt neuen „Zugang“ zur NCBI-Datenbank, holt noch keine Daten

Bio::DB::GenBank (2)

Zugriff auf einzelne Sequenzen

- ▶ mit bekanntem Namen (ID)
- ▶ mit bekannter GI-Nummer
- ▶ mit bekannter Accession-Number
- ▶ ...

Methoden:

```
$gb->get_Seq_by_id(<id>);  
$gb->get_Seq_by_gi(<gi>);  
$gb->get_Seq_by_acc(<accession>);  
...
```

... geben Bio::SeqI-Objekte zurück

Bio::DB::Query::GenBank

- ▶ Klasse für Anfragen an die NCBI GenBank
- ▶ wiederum exemplarisch

Konstruktor:

```
$query = Bio::DB::Query::GenBank->new();
```

Argumente:

- ▶ -db : 'protein' oder 'nucleotide'
- ▶ -query : die Anfrage (wie man sie im Suchfeld bei NCBI eingeben würde)
- ▶ -mindate : minimales Einstelldatum
- ▶ -maxdate : maximales Einstelldatum

Alternativ:

- ▶ -ids : Liste von GIs, überschreibt query

Query kann in `Bio::DB::GenBank::get_Stream_by_query()` benutzt werden.

Bio::DB::Query::GenBank (2)

Methoden:

```
$query->count(); #Anzahl der Ergebnisse der Anfrage
```

```
$query->ids(); #Liste von Gls als Ergebnis der Anfrage
```

Die Gls können in den Methoden von Bio::DB::GenBank benutzt werden, z.B.

```
$query = Bio::DB::Query::GenBank->new(-db => 'protein',  
                                     -query => 'coli_chromosome');
```

```
@ids = $query->ids();
```

```
$gb->Bio::DB::GenBank->new();
```

```
for $id (@ids){  
    $gb->get_Seq_by_gi($i);  
}
```

Bio::DB::GenBank (revisited)

Zugriff auf Mengen von Sequenzen

```
$gb->get_Stream_by_query(<query>); #Query-Objekte von letzter Folie
```

```
$gb->get_Stream_by_gi([<gi1>,<gi2>,...]);
```

```
$gb->get_Stream_by_acc([<acc1>,<acc2>,...]);
```

...

... geben Bio::SeqIO-Objekte zurück,

Zugriff auf einzelne Sequenzen mit next_seq()

Bio::SeqFeatureI

- ▶ Interface für Sequenzfeatures (Coding Sequences, repetitive Elemente,...)
- ▶ Objekte werden von Bio::Seq::get_SeqFeatures() zurückgegeben

Methoden:

```
$f->primary_tag(); #Tag des Features (z.B. 'CDS', 'exon', ...)  
$f->display_id(); #Name des Features zur Anzeige  
$f->seq(); #Sequenz(abschnitt) fuer dieses Feature (Bio::PrimarySeq)  
$f->spliced_seq() #ggf. gespleisste Sequenz, sonst wahrsch. Originalseq.  
$f->location() #Ort auf der Sequenz (Bio::LocationI, naechste Folie)  
$f->get_SeqFeatures() #Liste von Subfeatures (wenn vorhanden)  
$f->strand() #1 fuer den aktuellen und -1 fuer den Gegenstrang  
$f->annotation() #Annotationen (uebernaechste Folie)
```

Bio::LocationI

- ▶ Interface für Start- und Endpositionen auf einer Sequenz
- ▶ Objekte werden von Bio::SeqFeature::location () zurückgegeben

Methoden:

`$loc->start()` *#Starposition auf der Sequenz*

`$loc->end()` *#Endposition auf der Sequenz*

... und weitere Methoden in abgeleiteten Klassen.

Methoden in abgeleiteten Klassen können deutlich mehr Eigenschaften beinhalten!

Bio::AnnotationCollection

- ▶ Annotationen zu einem Feature
- ▶ Objekte von Bio::SeqFeatureI::annotation() zurückgegeben

Methoden:

`$ac->get_num_of_annotations()` *#Anzahl der Annotationen*

`$ac->get_all_annotation_keys()` *#Schlüssel (Namen) aller Annotationen*

`$ac->get_Annotations(<key>)` *#Liste aller Annotation zu einem Schlüssel*

Annotationsobjekte → nächste Folie

Bio::AnnotationI

- ▶ Interface für Annotationen (Genzugehörigkeit, Anmerkungen, ähnliche Gene,...)
- ▶ Objekte von Bio::AnnotationCollectionI :: get_Annotations() und Bio::SeqFeatureI :: get_Annotations() zurückgegeben

Methoden:

`$annot->tagname()` *#Tag der Annotation (entspricht key)*

`$annot->as_text()` *#textuelle Beschreibung des Wertes*