

# Sviluppo e-commerce avifauna.fem2ambiente.com

**Mattia Curatitoli**

Relatore: **Prof. Daniela Micucci**

Correlatore: **Dr. Emanuele Ferri**

Ottobre 2015

## FEM2 - Ambiente

Uno spin-off del Dipartimento di Biotecnologie e Bioscienze  
dell'Università degli studi di Milano-Bicocca

La mission dell'azienda è creare prodotti e servizi per il largo  
pubblico finalizzati alla conoscenza e tutela della biodiversità.

# Ensembl e accesso ai dati

Ensembl è un progetto scientifico dell'EBI

L'obiettivo del progetto è quello di fornire ai propri utenti accesso a una vasta collezione di **dati genetici**

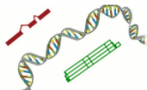
Il progetto supporta 70 specie animali, di cui vengono fornite le sequenze genomiche, set dei geni associati, allineamenti su larga scala e tra multiple specie, dati su variazioni e regolazione genica

Il progetto supporta 70 specie animali, di cui vengono fornite le sequenze genomiche, set dei geni associati, allineamenti su larga scala e tra multiple specie, dati su variazioni e regolazione genica

La mole complessiva dei dati messi a disposizione ammonta a diverse migliaia di GigaBytes

## Core

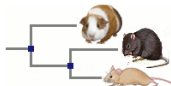
Sequence, genes and other [automated annotation](#)



- [Perl API](#)
- [Database schema](#)
- [Tutorial](#)

## Comparative genomics

Homologues, paralogues and protein families



- [Perl API](#)
- [Database schema](#)
- [Tutorial](#)

## Variation

SNPs, somatic mutations and structural variants



- [Perl API](#)
- [Database schema](#)
- [Tutorial](#)

## Regulation

Regulatory features, motifs and oligoprobes



- [Perl API](#)
- [Database schema](#)
- [Tutorial](#)

I dati del progetto possono essere acceduti in vari modi



I dati del progetto possono essere acceduti in vari modi

### Modalità “manuali”

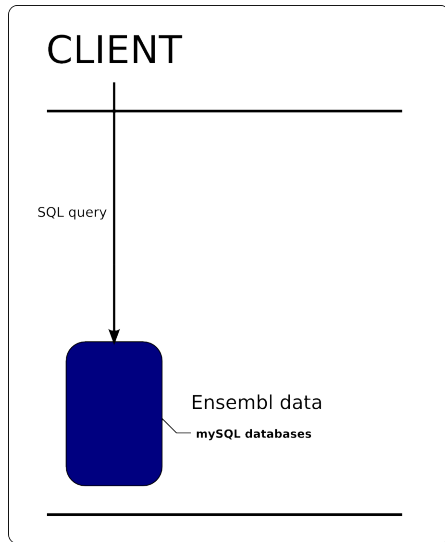
- Genome browser
- Ensembl Biomart

### Modalità programmatiche

- client SQL
- Perl API
- REST API (beta)

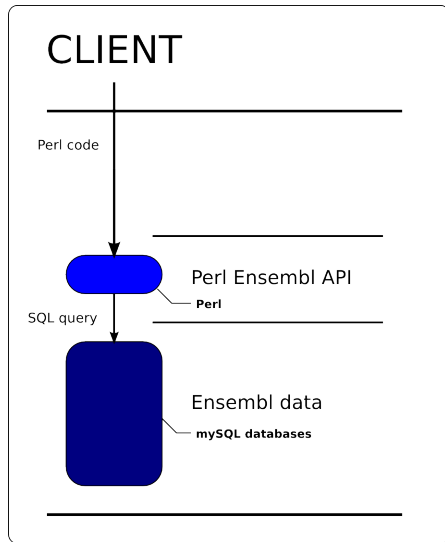
# Accesso programmatico

Accesso diretto ai database  
mySQL pubblici del progetto  
tramite client SQL



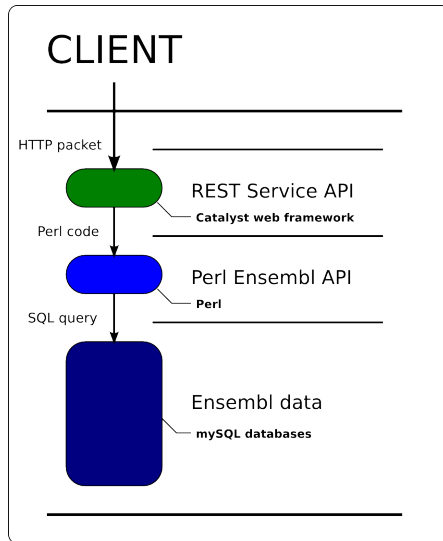
# Accesso programmatico

Accesso ai database tramite  
l'API Perl



# Accesso programmatico

Accesso ai database tramite  
l'API REST



Nel corso degli anni sono state sviluppate diverse **librerie alternative** che forniscono funzionalità simili a quelle del modulo ufficiale Perl

Tali librerie, sviluppate da terze parti come progetti open-source, necessitano di **frequenti aggiornamenti** per continuare a funzionare con le release Ensembl più recenti

L'API REST è stata rilasciata nel Settembre 2012, ed è attualmente in fase beta

Il servizio consente l'accesso programmatico **language-agnostic** ai dati del progetto Ensembl

I dati sono organizzati in endpoints,  
l'interrogazione del database avviene via HTTP.  
La query è codificata nell'URL

I dati sono organizzati in endpoints,  
l'interrogazione del database avviene via HTTP.  
La query è codificata nell'URL

`beta.rest.ensembl.org +`



I dati sono organizzati in endpoints,  
l'interrogazione del database avviene via HTTP.  
La query è codificata nell'URL

```
beta.rest.ensembl.org +  
/taxonomy/id +
```

I dati sono organizzati in endpoints,  
l'interrogazione del database avviene via HTTP.  
La query è codificata nell'URL

```
beta.rest.ensembl.org +  
/taxonomy/id +  
/Homo sapiens?content-type=application/xml
```

Il servizio REST è più semplice da interfacciare,  
ma richiede in ogni caso

Il servizio REST è più semplice da interfacciare,  
ma richiede in ogni caso

- la gestione della connessione HTTP
- la costruzione dell'URL corrispondente alla query
- il parsing del pacchetto HTTP di risposta
- l'estrazione dei dati ritornati

Il servizio REST è più semplice da interfacciare,  
ma richiede in ogni caso

- la gestione della connessione HTTP
- la costruzione dell'URL corrispondente alla query
- il parsing del pacchetto HTTP di risposta
- l'estrazione dei dati ritornati

ensembl-rest si occupa di questi dettagli, nascondendoli all'utente

# La libreria

Ogni endpoint del servizio ha un metodo Ruby corrispondente.

Ogni endpoint del servizio ha un metodo Ruby corrispondente.

```
beta.rest.ensembl.org/taxonomy/id
```



```
Taxonomy.taxonomy_id
```



Un esempio di interrogazione:

```
require 'bio-ensembl-rest'  
include EnsemblRest  
  
EnsemblRest.connect_db  
Taxonomy.taxonomy_id 'Homo sapiens'
```

```
{  
  "scientific_name" : "Homo sapiens",  
  "parent" : {  
    "scientific_name" : "Homo",  
    "name" : "Homo",  
    "id" : "9605",  
    "leaf" : 0,  
    "tags" : {  
      "scientific name" : ["Homo"],  
      "name" : ["Homo"],  
      "authority" : ["Homo Linnaeus, 1758"]  
    }  
  },  
  ...  
}
```

ensembl-rest è scritta in Ruby



Il codice sorgente è gestito  
con git e ospitato su GitHub



ensembl-rest utilizza Travis CI  
per la *continuous integration*  
del codice

## Ensembl Rest

A Ruby library for the RESTful Ensembl API.

build **passing**

## ensembl-rest è integrata con BioRuby



### BioRuby

Open source bioinformatics library for Ruby

#### Announcements

- Biogem is community developed plugins for the BioRuby. See the [complete list](#), [generic information](#), [README](#), [HOWTO](#) and [Tutorial](#).

#### Overview

- BioRuby comes with a comprehensive set of free development tools and libraries for bioinformatics and molecular biology, for the Ruby programming language. BioRuby has components for sequence analysis, pathway analysis, protein modelling and phylogenetic analysis; it supports many widely used data formats and provides easy access to databases, external programs and public web services, including BLAST, KEGG, GenBank, MEDLINE and GO.
- BioRuby comes with a tutorial, documentation and an interactive environment, which can be used in the shell, and in the web browser.
- BioRuby is supported by the [Open Bioinformatics Foundation](#) or O|B|F.

#### Reference

BioRuby: Bioinformatics software for the Ruby programming language  
Naohisa Goto, Pjotr Prins, Mitsuteru Nakao, Raoul Bonnal, Jan Aerts and Toshiaki Katayama  
*Bioinformatics* (2010) 26(20): 2617-2619. doi: 10.1093/bioinformatics/btq475



#### Documentation

- [Wiki page](#)



#### Download

- [RubyGems package](#)

ed è distribuita come biogem (plugin BioRuby)

84	 <b>bio img metadata</b>	Reads metadata from integrated microbial genomes (img) metadata (...)	Ben J. Woodcroft	0.0.1	8 months	...			456	123	0	0	
85	 <b>bio qtlHD</b>	(...) pre		pre					260	13			
86	 <b>bio gggenome</b>	Gggenome Ruby client (...) new	Mitsuteru Nakao	0.2.0	4 weeks	1				245	245	0	13
87	 <b>bio stockholm</b>	Parse stockholm sequence alignment format (...)	Ben J. Woodcroft	0.0.1	2 months					218	218		
88	 <b>bio statsample timeseries</b>	Timeseries modules for statsample (...)	Ankur Goel, Claudio Bustos	0.1.2	1 week	2		1 		213	213	3	67
89	 <b>bio raxml</b>	Ruby wrapper to run standard-raxml for maximum likelihood (...)	Fernando Izquierdo	0.1.0	8 weeks					169	169		
90	 <b>bio ensembl rest</b>	Ruby ensembl rest api (...) new	Alberto Donizetti	0.1.0	2 weeks	2				108	108	0	55

\*) gem **downloads** (total & last 90 days) from [rubygems.org](http://rubygems.org) as of 2013-09-13  
\*\*) number of commits on [GitHub](https://github.com) (last 7 days & last 90 days)

La libreria è stata sviluppata nel contesto di un lavoro di stage presso il **Parco Tecnologico Padano** di Lodi