

Introduction to Bioinformatics

5th practice

Databases

General useful information

Databases are basic tools of bioinformatics, but be careful!!! The data is not reliable only because "it is in the database". Databases are incomplete and may contain errors.

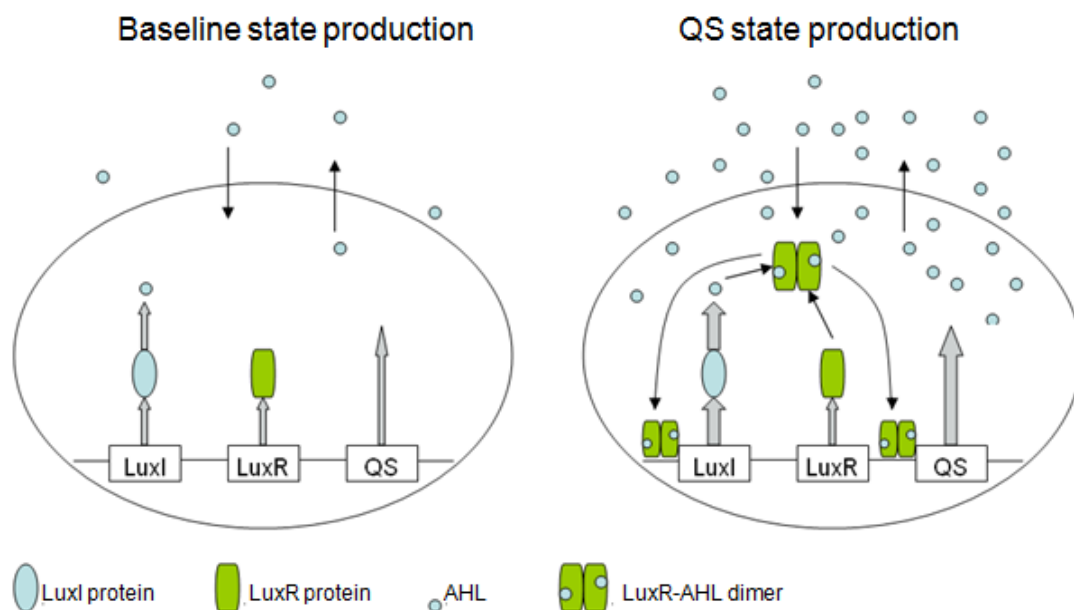
Many databases have their own data format and ID. Conversion is not always simple. Pay attention to what you are working with. Protein and even atom names may be very different for the same protein/atom.

Databases are updated, so save the files you work with before they're gone!

Brief biological background on LuxR and Quorum sensing

One way of bacterial cooperation is **quorum sensing**. QS is a form of gene regulation based on cell-density, which depends on inter-cellular communication involving the production and response to signalling molecules. Using quorum sensing means the ability to achieve coordinated behaviour, which ensures competitive advantage to a microbial population. A colony can adapt to the environment with this mechanism, can use the resources more efficiently and can defend itself against harmful microorganisms. For pathogenic bacteria it often plays a role in activating the mechanism that triggers infection of the host.

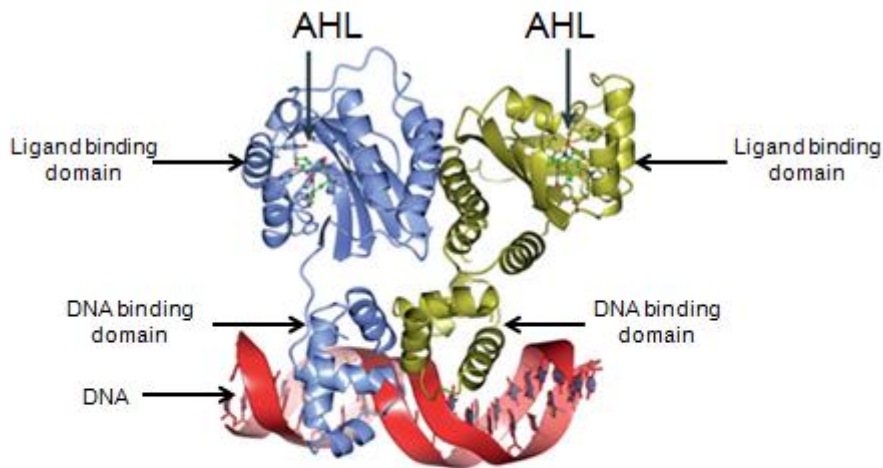
There are several types of QS systems. In Gram-negative bacteria the most common are the **LuxI/R systems**. These contain **regulator (R)** and **synthesis (I) genes**. Having a system with one regulator and one synthesis gene, the scheme of the process is described below.



The synthesis gene has a ground state activity, produces **AHL** (acyl-homoserine lacton) molecules which will be distributed in the extracellular and intracellular space by diffusion through the cell wall and membrane. These small molecules contain a homoserine-lacton

ring and some kind of acyl functional group. The variability of acyl groups ensures that different QS systems can work separately in the same environment or even in the same cell.

When the density of AHL producing cells increases, the intracellular concentration of AHL grows. The **luxR protein only can bind the AHL in high concentration**, and after binding it two luxR proteins form a **dimer structure**, which **binds to the DNA** and has **regulator functions**.



Crystal structure of a DNA binding luxR transcription regulator protein dimer:

It has two main parts:

the ligand binding domains bind AHL molecules with suitable structure, and the DNA binding parts connect with the major groove of the DNA double helix with their helix-turn-helix (HTH) domain.

http://www.nature.com/nrmicro/journal/v14/n9/fig_tab/nrmicro.2016.89_F2.html

It has two roles: at first, it **upregulates the synthesis gene** and **autocatalyses the regulator gene** (producing regulator protein). The system achieves a switch-like behaviour with this positive feedback loop. The second role of the regulator protein is **activating the factor producing genes**. Factors are the so called “public goods”, the materials that are secreted into the environment by the cells, for example enzymes. The production of these secreted factors is really energy consuming and one cell can produce only much smaller amount than the optimal level. So, if one or just some cells start to secrete the factors, they die. They run out of energy and do not achieve the goal of the process, for example, dissolve the food, because the factor concentration could not be enough. The members of a bacterial colony start the production and secretion of factors when they reach the efficient colony density to make enough factors for changing the environment in a required way. It means that the **density information comes from the concentration of signal and luxR proteins can sense this density, they do counting for the cells**.

Example and exercise

On today's session we will get familiar with a few databases while finding useful information about LuxR transcriptional regulator. Feel free to look around on the sites, there are a lot of useful links, tools categorised information that we will not have time to mention.

NCBI

National Center for Biotechnology Information (NCBI) is part of the United States National Library of Medicine (NLM), a branch of the National Institutes of Health. It is a collection of databases and tools.

Go to: <http://www.ncbi.nlm.nih.gov/>

The screenshot shows the NCBI homepage with a navigation menu on the left, a search bar at the top, and a main content area with several sections: 'Welcome to NCBI', 'Popular Resources', and 'NCBI Announcements'. The 'Welcome to NCBI' section includes a description of the center's mission and links to 'About the NCBI', 'Mission', 'Organization', 'NCBI News', and 'Blog'. Below this are six main activity icons: Submit (Deposit data or manuscripts into NCBI databases), Download (Transfer NCBI data to your computer), Learn (Find help documents, attend a class or watch a tutorial), Develop (Use NCBI APIs and code libraries to build applications), Analyze (Identify an NCBI tool for your data analysis task), and Research (Explore NCBI research and collaborative projects). The 'Popular Resources' section lists items like PubMed, Bookshelf, PubMed Central, PubMed Health, BLAST, Nucleotide, Genome, SNP, Gene, Protein, and PubChem. The 'NCBI Announcements' section includes news about the CCDS release 20 for human now public in Gene (dated 04 Oct 2016) and the October 11th NCBI Minute: BLAST+ 2.5.0 with Support for HTTPS, accession.version Identifiers and Much...

If we want to execute a specific job, we can choose it from the list on the left side (DNA&RNA, Genes&Expression, ...). On the right side we can find the the most popular resources of NCBI. On the top of the page we can search for a specific phrase (gene, protein, person,...).

Here search for LuxR.

The screenshot shows the NCBI search results page for the query 'LuxR'. The search bar at the top contains 'LuxR' and a 'Search' button. Below the search bar, it states 'Results found in 22 databases for "LuxR"'. The results are organized into several categories, each with a table of database names, hit counts, and brief descriptions.

Category	Database	Count	Description
Literature	Books	24	books and reports
	MeSH	8	ontology used for PubMed indexing
	NLM Catalog	0	books, journals and more in the NLM Collections
	PubMed	821	scientific & medical abstracts/citations
	PubMed Central	2,841	full-text journal articles
Health	ClinVar	0	human variations of clinical significance
	dbGaP	0	genotype/phenotype interaction studies
	GTR	0	genetic testing registry
	MedGen	0	medical genetics literature and links
	OMIM	0	online mendelian inheritance in man
	PubMed Health	0	clinical effectiveness, disease and drug reports
	UniGene	0	clusters of expressed transcripts
Genomes	Assembly	1	genome assembly information
	BioProject	18	biological projects providing data to NCBI
	BioSample	0	descriptions of biological source materials
	Clone	0	genomic and cDNA clones
	dbVar	0	genome structural variation studies
	Genome	6	genome sequencing projects by organism
Genes	EST	22	expressed sequence tag sequences
	Gene	17,281	collected information about gene loci
	GEO DataSets	56	functional genomics studies
	GEO Profiles	272	gene expression and molecular abundance profiles
	HomoloGene	0	homologous gene sets for selected organisms
	PopSet	55	sequence sets from phylogenetic and population studies
Proteins	Conserved Domains	12	conserved protein domains
	Protein	692,374	protein sequences
	Protein Clusters	1,682	sequence similarity-based protein clusters
	Structure	62	experimentally-determined biomolecular structures
Chemicals	BioSystems	2,199	molecular pathways with links to genes, proteins and chemicals
	PubChem BioAssay	100	bioactivity screening studies
	PubChem Compound	0	chemical information with structures, information and links
	PubChem Substance	20	deposited substance and chemical information

You can see the databases and the number of hits categorized by the type of database. We will first look at PubMed, so **open the PubMed link in a new tab.**

The screenshot shows the PubMed website interface. At the top, there is a search bar with 'LuxR' entered. Below the search bar, there are options for 'Format: Summary', 'Sort by: Most Recent', and 'Send to'. A box highlights search results: 'See 2 articles about luxR gene function. See also: luxR LuxR family transcriptional regulator in the Gene database. luxR in Vibrio fischeri ES114 Agrobacterium fabrum str. C58 Sinorhizobium sp. M14 All 3 Gene records'. The main search results section shows two items: 1. 'Advances in the research of LuxR family protein in quorum-sensing system of gram-negative bacteria' by Chen Z, Xiang J, et al. (2016). 2. 'Flavones as Quorum Sensing Inhibitors Identified by a Newly Optimized Screening Platform Using Chromobacterium violaceum as Reporter Bacteria' by Skogman ME, Kanerva S, et al. (2016). On the right side, there is a 'Results by year' bar chart and a 'Related searches' section with terms like 'luxR family transcriptional regulator', 'quorum sensing luxR', 'luxR streptomyces', and 'luxI luxR'.

PubMed is a collection of biomedical papers. The abstract of a biomedical paper is always readable. We can execute a quick search or setup an advanced search with many different fields (title, author, publication date, issue, journal, language, ...)

**How many scientific article contains the expression “luxR” in its title or abstract?
 How many of them are from 2016, 2015, 2014, 2013, 2012, 2011, 2010?
 When was the first mention of luxR?**

Notice the URL keeps it simple. This is very helpful in reaching data via script or commands (not only on this site). Also, see the YouTube tutorial link on the advanced search site in the right hand upper corner? (2:26 minutes)

“scholar.google.com” is also frequently used for finding papers you are looking for.

Nucleotide is the database of the nucleotide sequences. It contains the **GenBank** database, which is an open access, annotated collection of all publicly available nucleotide sequences and their protein translations. GenBank (USA) exchanges data on a daily basis with EMBL (Europe) and DDBJ (Japan) databases.

Hits of Nucleotide are GenBank records that contain the sequences and annotation. The files describe a gene, a gene group, a chromosome or a full genome. GenBank refers to the database and the format.

Go back to the search results tab. Find the Nucleotide database and click...

Genomes

Assembly	1	genome assembly information
BioProject	18	biological projects providing data to NCE
BioSample	0	descriptions of biological source materia
Clone	0	genomic and cDNA clones
dbVar	0	genome structural variation studies
Genome	6	genome sequencing projects by organis
GSS	9	genome survey sequences
Nucleotide	764,193	DNA and RNA sequences
Probe	0	sequence-based probes and primers

Here you can (almost) see that Nucleotide contains all kinds of nucleotide sequences: partial or full genes, cds (coding DNA sequence), full gene, mRNA, ... With the help of the links on the left, you can narrow your search results.

How many genomic bacterial sequences are in the database?

Species: Animals (120), Plants (157), Fungi (33), Protists (112), Bacteria (763,493), Archaea (58), Viruses (82), Customize ...

Molecule types: genomic DNA/RNA (764,168), mRNA (76), Customize ...

Source databases: INSDC (GenBank) (352,528), RefSeq (411,767), Customize ...

Genetic compartments: Chloroplast (28), Plasmid (1,620), Plastid (73)

Sequence length: Custom range...

Release date: Custom range...

Revision date: Custom range...

Clear all

Summary 20 per page Sort by Default order Send to: Filters: Manage Filters

See [luxR LuxR family transcriptional regulator](#) in the Gene database
luxR reference sequences [Protein \(1\)](#)

Items: 1 to 20 of 764307

Found 764338 nucleotide sequences. Nucleotide (764307) EST (22) GSS (9)

- [Vibrio harveyi LuxR \(luxR\) gene, partial cds](#)
916 bp linear DNA
Accession: DQ108980.1 GI: 74418827
[GenBank](#) [FASTA](#) [Graphics](#)
- [Sinorhizobium meliloti strain AK21 acyl-homoserine lactone syntetase, hypothetical protein, and LuxR \(luxR\) genes, complete cds, and hypothetical protein gene, partial cds](#)
3,162 bp linear DNA
Accession: JX049333.1 GI: 401022230
[GenBank](#) [FASTA](#) [Graphics](#)
- [Vibrio fischeri strain SJ12 LuxR \(luxR\) and LuxI \(luxI\) genes, partial cds](#)
1,125 bp linear DNA
Accession: AY292984.1 GI: 37223252
[GenBank](#) [FASTA](#) [Graphics](#) [PopSet](#)
- [Aliivibrio fischeri strain MJ1 LuxR \(luxR\) gene, partial cds, luxR-luxI intergenic spacer, complete sequence, and LuxI \(luxI\) gene, partial cds](#)
218 bp linear DNA
Accession: HQ436495.1 GI: 323650551
[GenBank](#) [FASTA](#) [Graphics](#) [PopSet](#)

Results by taxon

Top Organisms [\[Tree\]](#)

- Salmonella enterica (101427)
- Escherichia coli (90332)
- Staphylococcus aureus (59044)
- Mycobacterium tuberculosis (44420)
- Pseudomonas aeruginosa (39445)
- All other taxa (429639)

More ...

Find related data

Database: [Select](#)

Find items

Search details

LuxR[All Fields]

Search See more...

Recent activity

Turn Off Clear

Q LuxR (764307)

Hits have title, length, Accession and GI (Gene Info) listed. Records can be opened in different formats. **Open one of them in GenBank. Look around, look at the record. What kind of information can a GenBank record store? How long is your sequence (is it a full sequence or is it just partial)? What is the source organism?** In a GenBank record the annotation is on the top, the sequence starts with "ORIGIN" and ends with "//".

```

ACCESSION   DQ108980 S83266
VERSION    DQ108980.1 GI:74418827
  
```

Accession is stable, universal, and reportable. Version tracks changes in sequence (!!!) not in annotation. GI (Gene Info) is for NCBI internal use. More info on Accession format:

<https://www.ncbi.nlm.nih.gov/Sequin/acc.html>. Note that usage of GI number is not supported by NCBI anymore (as of 2016 september) and accession number is preferred. Nucleotide has all nucleotide sequences, which means it is highly redundant. The same gene's sequence can be uploaded from multiple sources. It contains sequences of varying quality, and contains both uncurated and curated data.

RefSeq (reference sequence database) is a subset of the NCBI database. It is manually curated, contains high quality sequence data, and is non-redundant; this means that each gene (or splice-form of a gene, in the case of eukaryotes), protein, or genome sequence is only represented once. Because of the high level of manual curation required, RefSeq does not cover all species, and is not comprehensive for the species that are covered so far. For easier comparison and analysis, RefSeq records have a consistent format and distinct accession series:

mRNAs and Proteins	
NM_123456	Curated mRNA
NP_123456	Curated Protein
NR_123456	Curated non-coding RNA
XM_123456	Predicted mRNA
XP_123456	Predicted Protein
XR_123456	Predicted non-coding RNA
Gene Records	
NG_123456	Reference Genomic Sequence
Chromosome	
NC_123455	Microbial replicons, organelle genomes, human chromosomes
AC_123455	Alternate assemblies
Assemblies	
NT_123456	Contig
NW_123456	WGS Supercontig

RefSeq and GenBank are different databases and have different accession formats. See at the nucleotide search results:

Source databases

INSDC (GenBank) (352,528)

RefSeq (411,767)

Customize ...

The NCBI **FTP site** can be used for bulk downloads, for example if you need all the Gram-negative bacteria sequence, all of the complete genomes, ect. **Go to** <ftp://ftp.ncbi.nlm.nih.gov/> . You can see genomes, refseq, pubmed,...

A few important file formats:

fna: full DNA (fasta format)

ffn: DNA decomposed to genes (fasta format)

faa: protein sequence decomposed to genes (fasta format)

gbk: EMBL type file (Genbank record - annotation information is added to the sequence)

ptt, gff: table with gene info

All these (and many more) databases are available online through the Entrez search engine. Entrez is NCBI's primary text search and retrieval system. Options include popular

configurable preset facet filters to help focus on specific kinds of results, an Advanced Search interface that facilitates constructing more sophisticated queries. Specialized search fields are available for each database and can be browsed and selected in the Search Builder section of the Advanced Search interface. Source and more info: <https://www.ncbi.nlm.nih.gov/books/NBK3837/>


Instead of carrying out searches of the NCBI database on the NCBI website, you can search directly from code, for example Python, R or even UNIX shell (this option is also through Entrez).

Another summarizing portal **ExpPASy** (easy to memorize :)) concentrates on proteins (at least that's what it's most used for). Go to <http://www.expasy.org/>.

The screenshot shows the ExpPASy Bioinformatics Resource Portal homepage. At the top left is the SIB logo and the ExpPASy logo. Below the logo is a search bar with the text "Query all databases" and a search button. On the left side, there is a navigation menu with sections: "Visual Guidance", "Categories" (with sub-items: proteomics, genomics, structure analysis, systems biology, evolutionary biology, population genetics, transcriptomics, biophysics, imaging, IT infrastructure, medicinal chemistry, glycomics), "Resources A..Z", and "Links/Documentation". The main content area features a "Featuring today" section with a card for "TMPred" (Predict trans-membrane regions in protein sequences) and a "How to use this portal?" section with a question mark icon and a list of bullet points: "Features and updates", "New to ExpPASy", and "Experienced ExpPASy users: what is different". On the right side, there are sections for "Popular resources" (UniProtKB, SWISS-MODEL, STRING, PROSITE) and "Latest News" (UniProtKB Knowledgebase release 2016_09 - 2016-10-05, Protein Spotlight: You want it darker - 2016-09-30).

ExpPASy contains databases and tools for proteomics, genomics and many more... It has a nice structure and lists many-many online tools. For example, you can see the **Categories** at the right hand side. **Select proteomics and then similarity search/alignment**. You can see the Databases on the left and Tools on the right. At Tools you can find ones we used on the last practice. Can you find T-Coffee?

The screenshot shows the "Databases" and "Tools" sections of the ExpPASy portal. The "Databases" section lists: "UniProtKB • functional information on proteins • [more]" and "MyHits • protein domains database and tools • [more]". The "Tools" section lists: "Alignment tools • Four tools for multiple alignments • [more]", "BLAST • sequence similarity search • [more]", "BLAST (UniProt) • BLAST search on the UniProt web site • [more]", "BLAST · NCBI • Biological sequence similarity search • [more]", "BLAST · PBIL • BLAST search on protein sequence databases • [more]", "Blast2Fasta • Blast to Fasta conversion • [more]", "boxshade • MSA pretty printer • [more]", "ClustalO (UniProt) • Align two or more protein sequences • [more]", and "ClustalW • Multiple sequence alignment • [more]".

We will use **UniProtKB**, so **click on**  **UniProtKB**. Next time you can go straight there: www.uniprot.org

Swiss-Prot (552,259)

Manually annotated and reviewed.

UniProtKB consists of two parts. [Swiss-Prot \(552,259\)](#) is the curated part (relatively

TrEMBL
(67,940,995)

Automatically annotated and not reviewed.

reliable) and [TrEMBL \(67,940,995\)](#) is translated from the nucleotide sequences of [EMBL](#), it is not reliable.

Note:



Type “LuxR” in the search interface on the top of the page.



UniProtKB results

Filter byⁱ

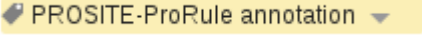
- Reviewed (296) Swiss-Prot
- Unreviewed (125,887) TrEMBL
- Popular organisms**
- E. coli K12 (23)
- B. subtilis (11)
- ALIFS (35)
- VIBHA (47)
- VIBFI (26)

- Function
- Names & Taxonomy
- Subcellular location
- Pathology & Biotech
- PTM / Processing
- Expression
- Interaction
- Structure
- Family & Domains
- Sequence
- Cross-references
- Entry information
- Miscellaneous
- Similar proteins

Entry	Entry name	Protein names	Gene names	Organism	Length
P12746	LUXR_ALIFS	Transcriptional activator protein L...	luxR	Aliivibrio fischeri (Vibrio fischeri)	250
P35327	LUXS_VIBF1	Transcriptional activator protein L...	luxR_VF_A0925	Vibrio fischeri (strain ATCC 700601 / ES114)	250
P21308	LUXR_VIBHA	HTH-type transcriptional regulator ...	luxR	Vibrio harveyi (Benecke harveyi)	205
O53720	O53720_MYCTU	LuxR family transcriptional regulat...	Rv0386, LH57_02070	Mycobacterium tuberculosis (strain ATCC 25618 / H37Rv)	1,085
Q8GBU2	Q8GBU2_VIBHA	LuxR	luxR	Vibrio harveyi (Benecke harveyi)	205
Q8NML3	RAMA_CORGL	HTH-type transcriptional activator ...	ramA_cg2831, Cgl2560	Corynebacterium glutamicum (strain ATCC 13032 / DSM 20300 / JCM 1318 / LMG 3730 / NCIMB 10025)	281
A0A050716.6	A0A050716.6_A0A050716.6	Bacterial	IsrR luxP_075413_05595	Arinobacter baumannii	238

Hits marked with are from Swiss-Prot. Notice the entry names consist of two parts: “NAME_ORGANISM”. The id and the entry names are unique. **Choose the one from *Aliivibrio fischeri* bacteria.**


With the help of this summary on the left, you can quickly get the information you want. **Look around on the page. What kind of information do you see? Try links, get familiar with the site.** Clicking on the names with little “i”-s, you can read summary information about it.

The flags  show the source of the information, so you can decide how much you believe it.

At “Family & Domains” you can see if there are annotated domains in the protein and where. **Clicking on the position you reach the domain in fasta format, ready to BLAST** (see on next practice).

What kind of domain(s) can be find in this protein?

Is a reliable 3D structure available for the entire protein?

At “Sequence” you can download it in fasta format. **Click on** . Notice the URL! With for example wget you can download more sequences at a time (if you know their id) without having to click your way through.

Entry history ⁱ	Integrated into UniProtKB/Swiss-Prot:	October 1, 1989
	Last sequence update:	February 1, 1994
	Last modified:	September 7, 2016
	This is version 100 of the entry and version 3 of the sequence. [Complete history]	

Remember? Databases are updated. Let’s say you start your research with searching the databases. By the time you write a paper on it, or you just want to go back and double check, it may be gone. It’s safer to save it (considering the amount of data of course).

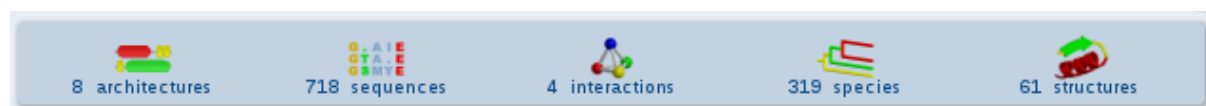
At “Similar proteins” you can find the links to the UniProt Reference Clusters.

Go back and find “Cross-references”. You can find links to many different types of databases. **Read them through**. (I don’t see any point of copying all of the titles, but they are important.)


Open the Pfam links in new tabs.


Pfam is a protein domain family database. There are two levels of quality to Pfam families. Pfam-A is based on manually annotated seed alignments (high quality) these ids start with PF and Pfam-B is automatically generated these ids start with PB.

In Pfam you can look at multiple alignments of sequences from multiple databases, view protein domain architectures, examine species distribution, follow links to other databases, view known protein structures.



Go through (read) the summary. Now we will look at the domain architectures. **Click on “Domain organisation” or on the architectures picture**. Examine the first hits. No wonder there were two links to Pfam. **Why? See the two binding parts? Which one is the HTH domain? How many proteins are there with similar architecture?** (There is a big difference in where you are looking at it from. HTH is probably more abundant and more frequently researched.)

There are 658 sequences with the following architecture: **Autoind_bind, GerE**
[Q216V3_RHOPB](#) [Rhodospseudomonas palustris (strain BisB18)] Transcriptional regulator, LuxR family {ECO:0000313|EMBL:ABD87583.1} (242 residues)

[Show](#) all sequences with this architecture.

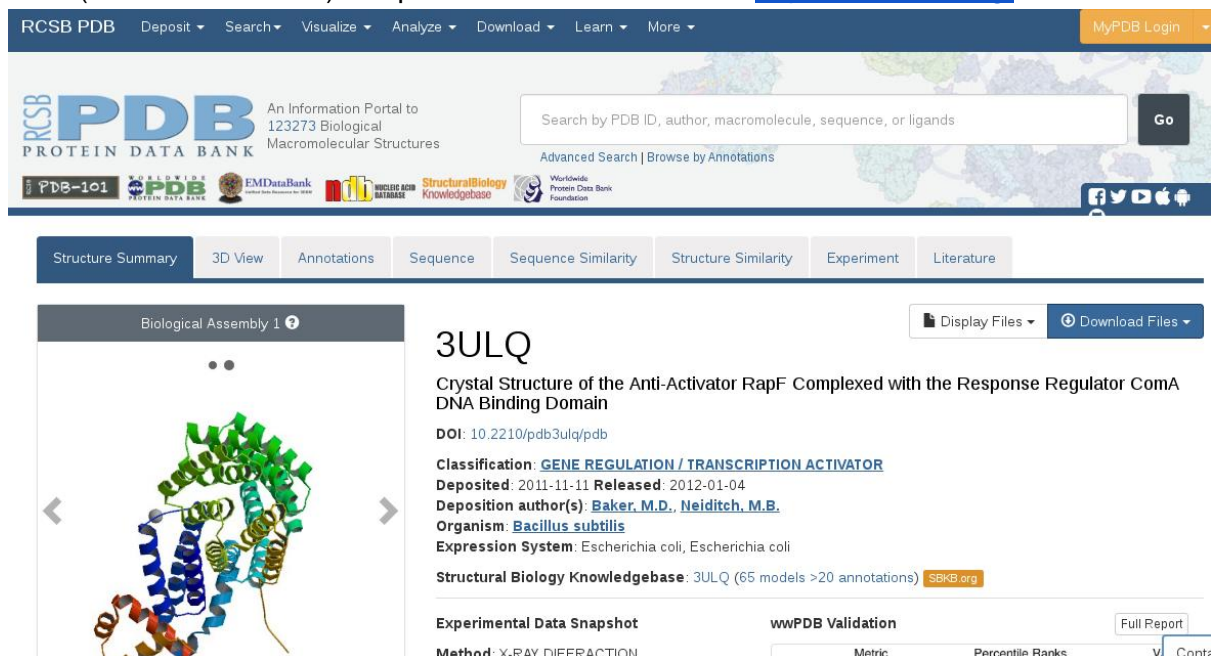
There are 13638 sequences with the following architecture: **Response_reg, GerE**
[W5Z0Q8_9ALTE](#) [Marinobacter salarius] LuxR family transcriptional regulator {ECO:0000313|EMBL:AH132078.1} (208 residues)

[Show](#) all sequences with this architecture.

Click on structures and select one by its PDB ID.

PDB ID	PDBe 3ulq
	RCSB PDB 3ulq
2KRF	PDBsum 3ulq
3ULQ	CATH 3ulq
3QP5	SCOP 3ulq
3QP5	ECOD 3ulq
3QP5	Proteopedia 3ulq
4WSZ	

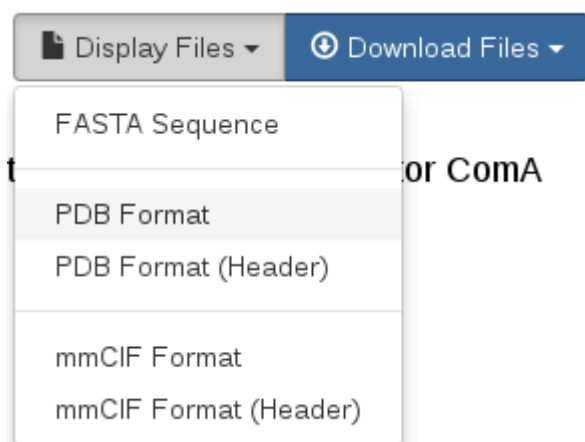
Click on the RCSB PDB link.

PDB (Protein Data Bank) is a protein structure database. <http://www.rcsb.org/>



Structures are defined with X-ray, NMR (nuclear magnetic resonance) or electron microscope. Records contain the 3D coordinates of biological molecules. IDs are 4 digits.

Look at the PDB Format. Starts with a “HEADER”, ends with “END”. The coordinates are noted “ATOM”. You can see how long and not concise it is. There are better formats, but everybody uses PDB and tools, applications are designed to use PDB.



Now we will look at how to browse genomes with **ENSEMBL**. **Go to**

<http://www.ensembl.org/>.

ENSEMBLE is a project of EBI and Sanger Institute. It is one of the largest genome search engines. Similar browsers can be found at NCBI and UCSC (University of California, Santa Cruz).

All genomes

-- Select a species --

Click on View full list of all Ensembl species. See all the species? Scroll to the bottom to Metazoa, Plants, Fungi,...

We need Bacteria, so **click on** [EnsemblBacteria](#).

[View full list of all Ensembl species](#)

Search for "lux". On the third page there are Lux-R hits.

MGMAQ_0952

Description	Lux R family transcriptional activator protein
Gene ID	MGMAQ_0952
Species	Magnetospira sp. QH-2
Location	Chromosome:946743-947519

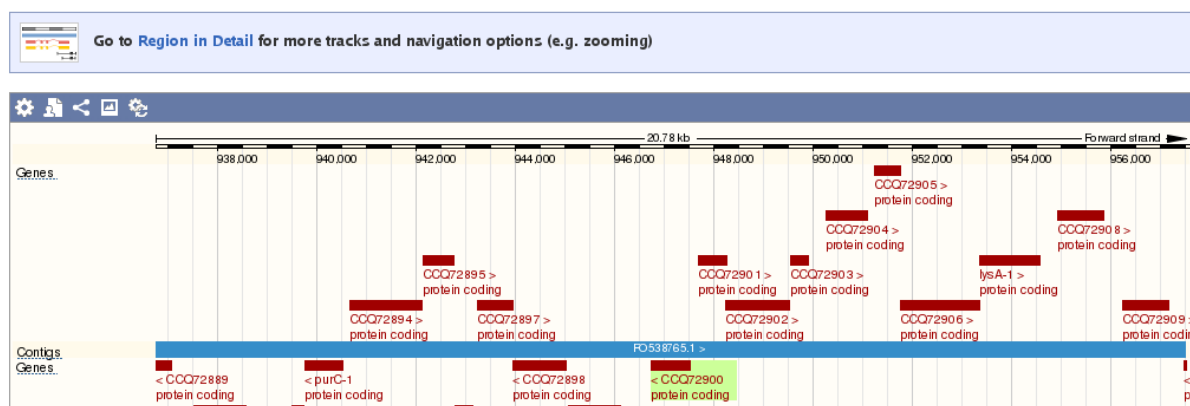
Selecting the MGMAQ_0952, you find summary information and the location of the gene.

Gene: MGMAQ_0952

Description	Lux R family transcriptional activator protein
Location	Chromosome: 946,743-947,519 reverse strand. ASM96813v1:FO538765.1
About this gene	This gene has 1 transcript (splice variant).
Transcripts	Show transcript table

Summary

Gene type	Protein coding
Annotation method	Protein coding genes annotated in ENA



Look at the

[List of species](#)

Since the Human genome is more complete and different than the Bacteria, we will have a quick look at that too. **Go back to the species list.**

http://www.ensembl.org/Homo_sapiens/Info/Index

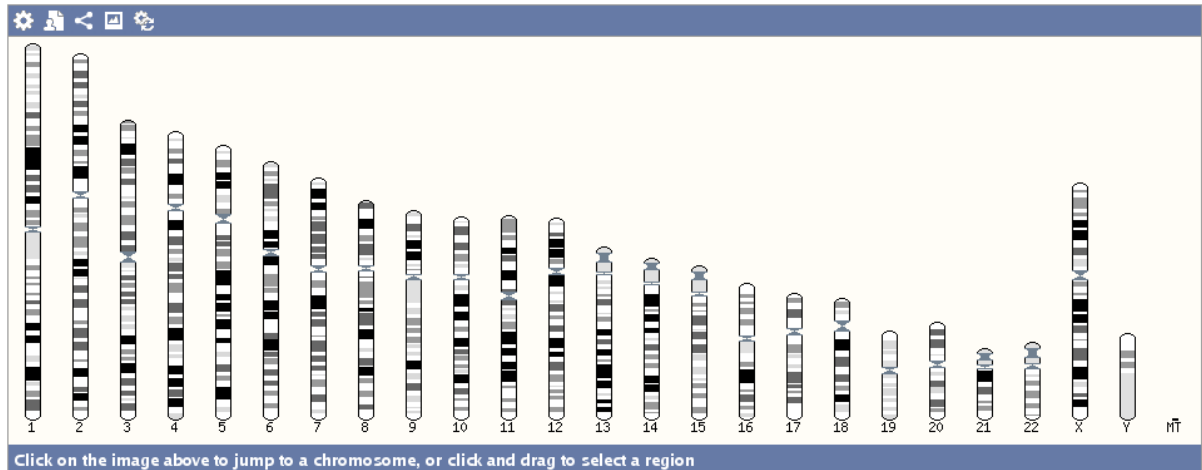


Human



View karyotype

Click on the link, then . Now you can see the banding patterns of all human chromosomes.



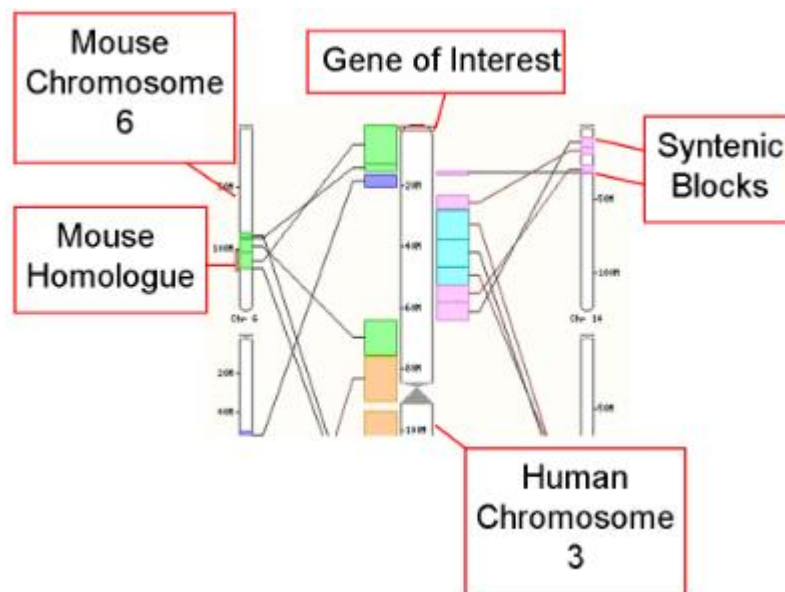
Other genome browsers

- UCSC
- NCBI
- Ensembl GRCh37

Checking the alternative browser can be very useful. All of these sites have their advantages and disadvantages.

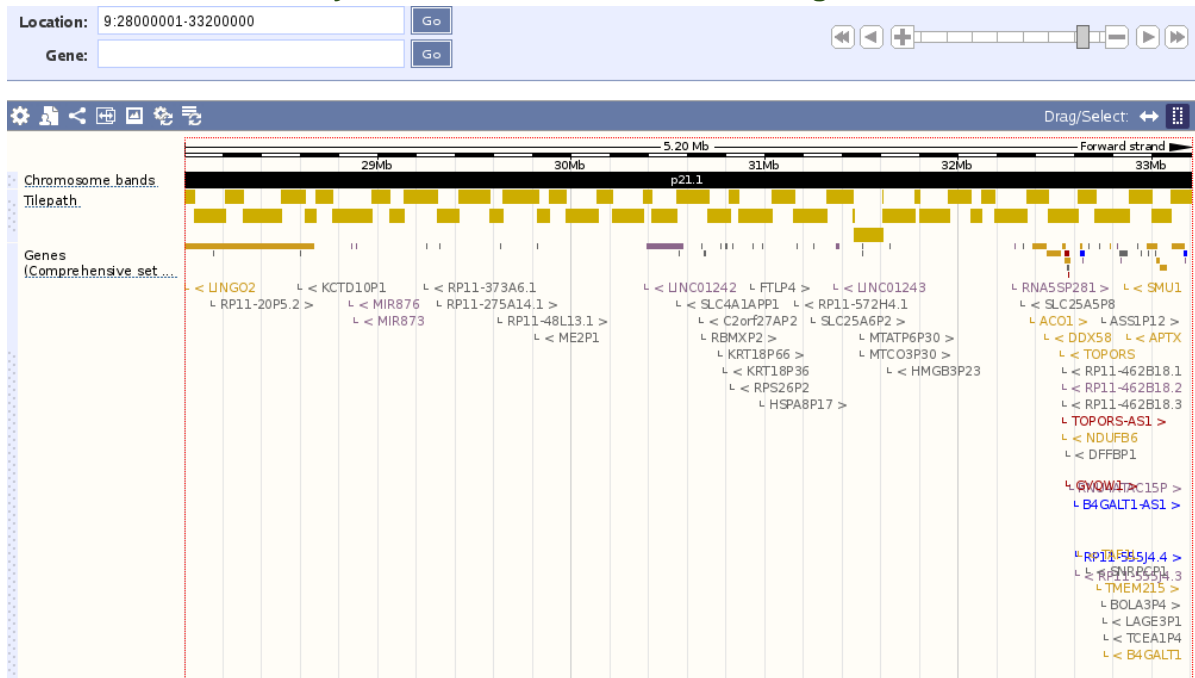
Select one of the chromosomes than chromosome summary.

See how dense some coding regions are? Look at the variations.

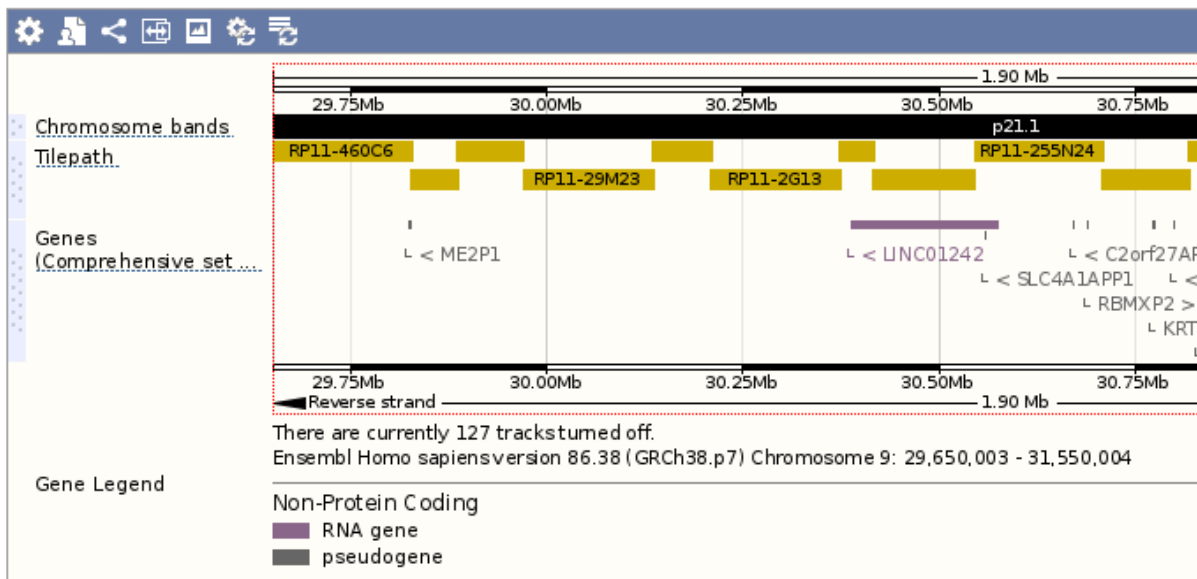


Click on **“Synten”**. This Shows the similar regions in different species. If you want more info click on the **“?”**.

Go back to the summary. Select a band and the offered region.



Here you can select the gene you are looking for, examine what other genes are in the area. With the scroll on the top right zoom in, to select a gene.



Check out the data you can see about the gene.

Selecting the gene tab,

30,388,935-30,575,012 **Gene: LINC01242** Transcript: LINC01242-001

Gene: LINC01242 ENSG00000280683

Description long intergenic non-protein coding RN
Location [Chromosome 9: 30,388,935-30,575,0](#)
GRCh38: CM000671.2

You can reach data on the gene level (earlier we were on the chromosome level). You can find orthologues/paralogues genes data or GO annotation (if given).