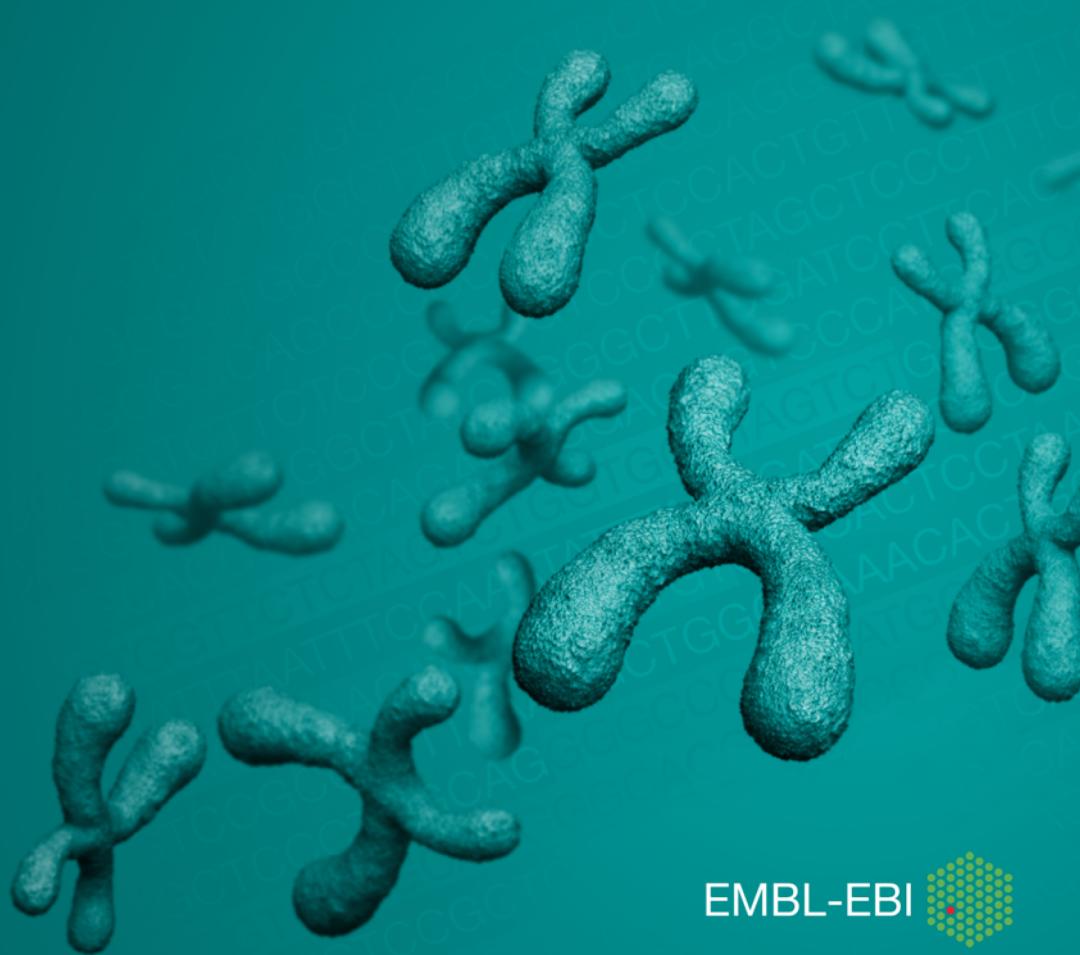


Metagenome Exchange (registry) Contextual data clearinghouse

Guy Cochrane
www.ebi.ac.uk



Two projects

- Metagenome Exchange (registry)
 - EMBL-EBI and ANL
 - BBSRC and NSF
- Contextual data clearinghouse
 - ELIXIR Implementation Study
 - Universities of Tromsø, Trento and Bari; MPI-Bremen; EMBL-EBI
- Imminent launch
- Starting June 2018

Relevance

- Participation, contribution, use
- Illustrate strategy to handle growth

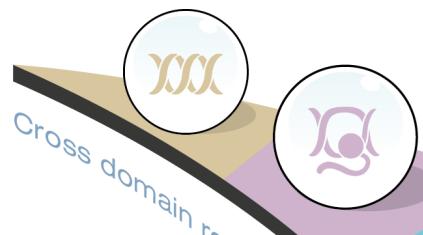
EMBL European Bioinformatics Institute

Genes, genomes & variation

European Nucleotide Archive
1000 Genomes

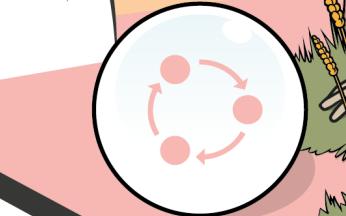
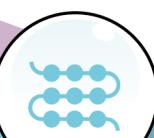
Ensembl
Ensembl Genomes

European Genome-phenome Archive
Metagenomics portal



Literature & ontologies

Europe PubMed Central
Gene Ontology
Experimental Factor
Ontology



Reactions, interactions & pathways

IntAct

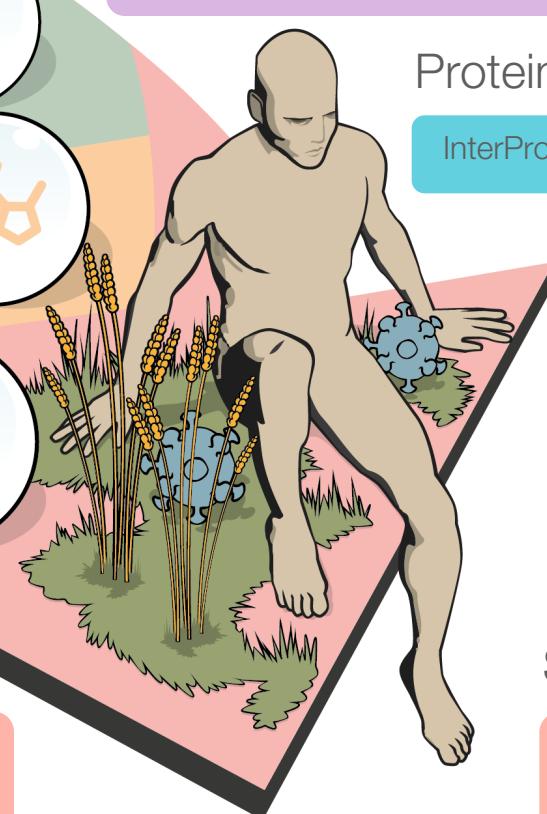
Reactome

MetaboLights

Gene, protein & metabolite expression

ArrayExpress
Expression Atlas

Metabolights
PRIDE



Protein sequences, families & motifs

InterPro

Pfam

UniProt

Molecular structures

Protein Data Bank in Europe
Electron Microscopy Data Bank

Chemical biology

ChEMBL

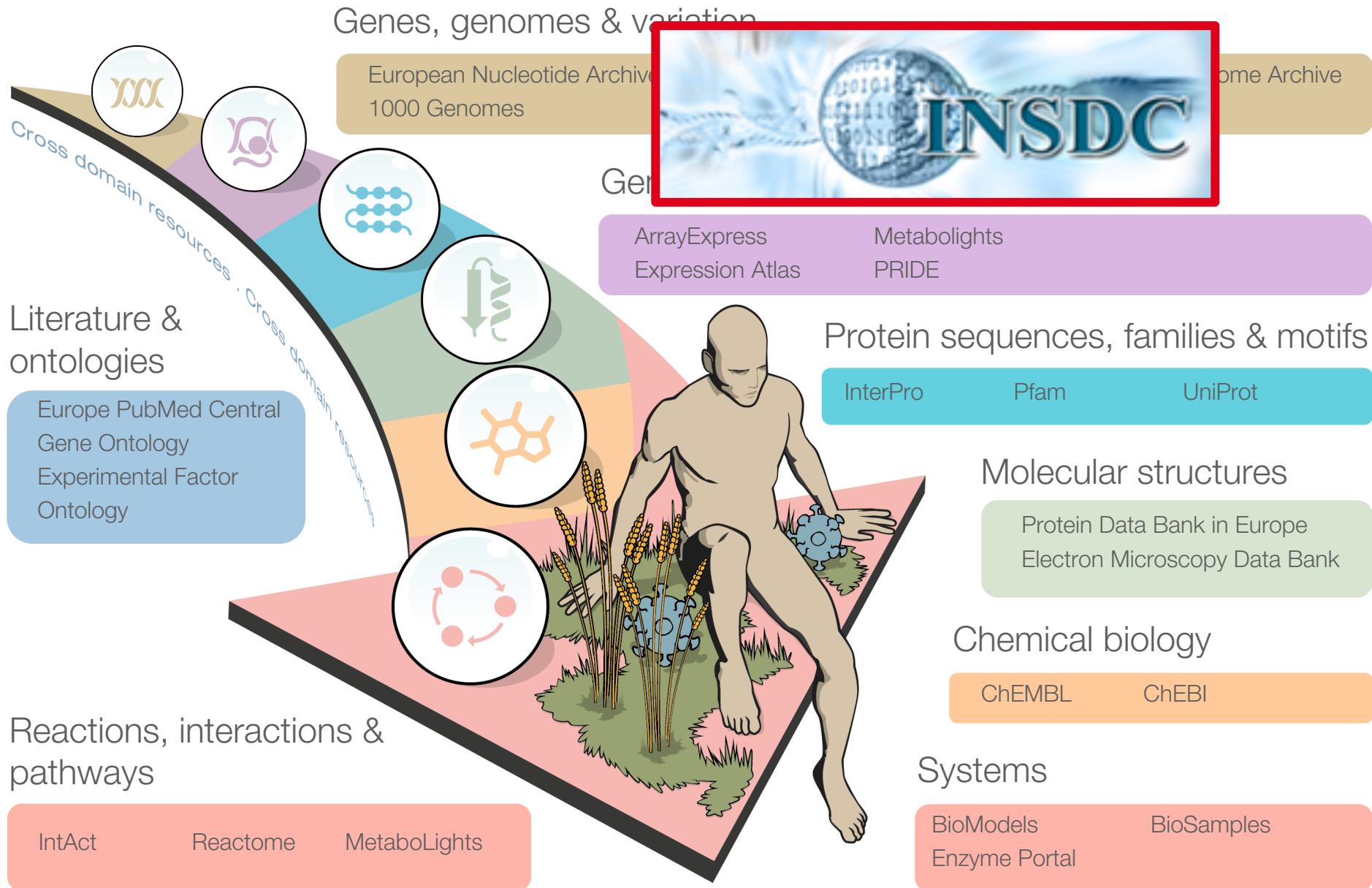
ChEBI

Systems

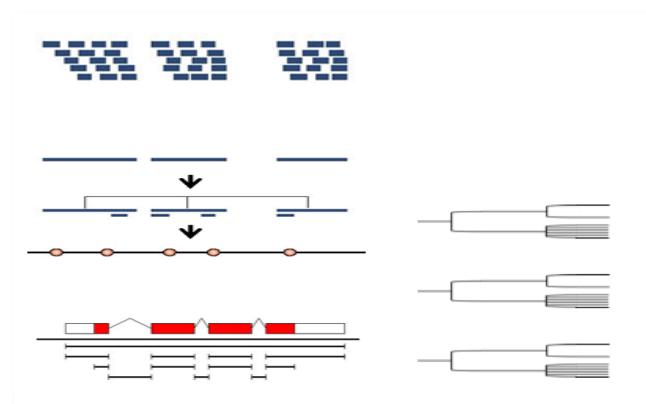
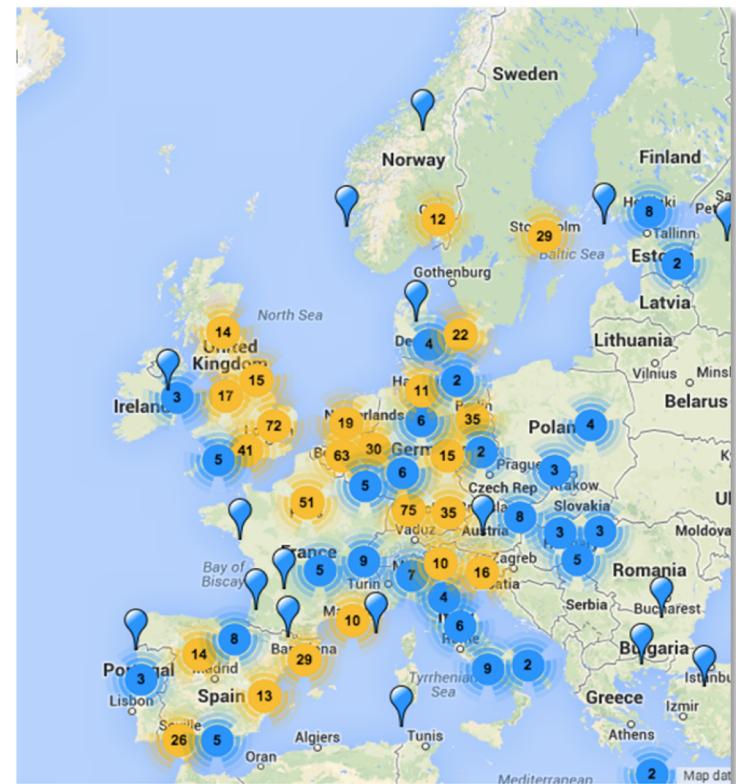
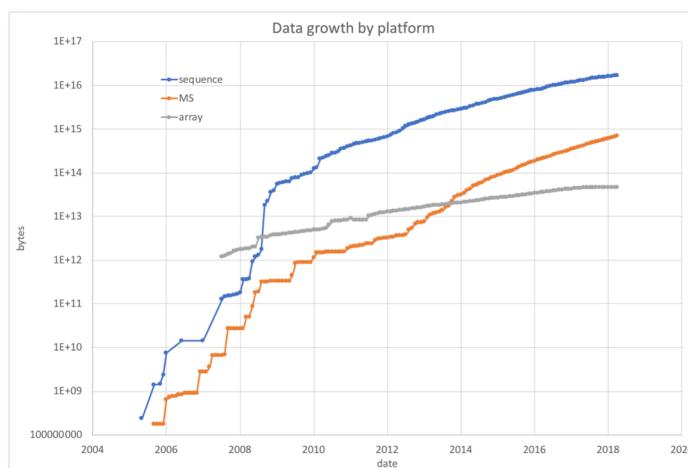
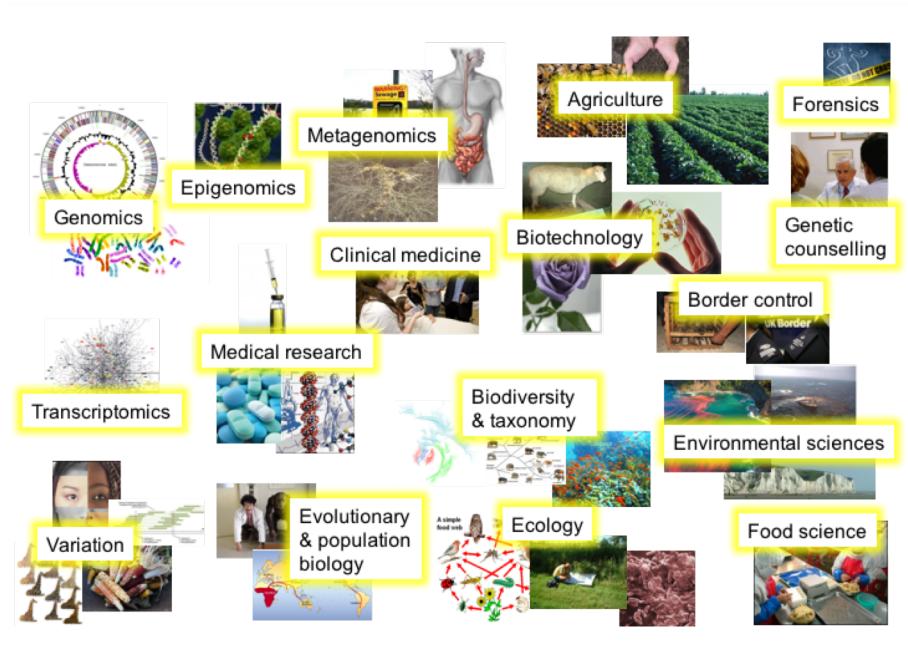
BioModels
Enzyme Portal

BioSamples

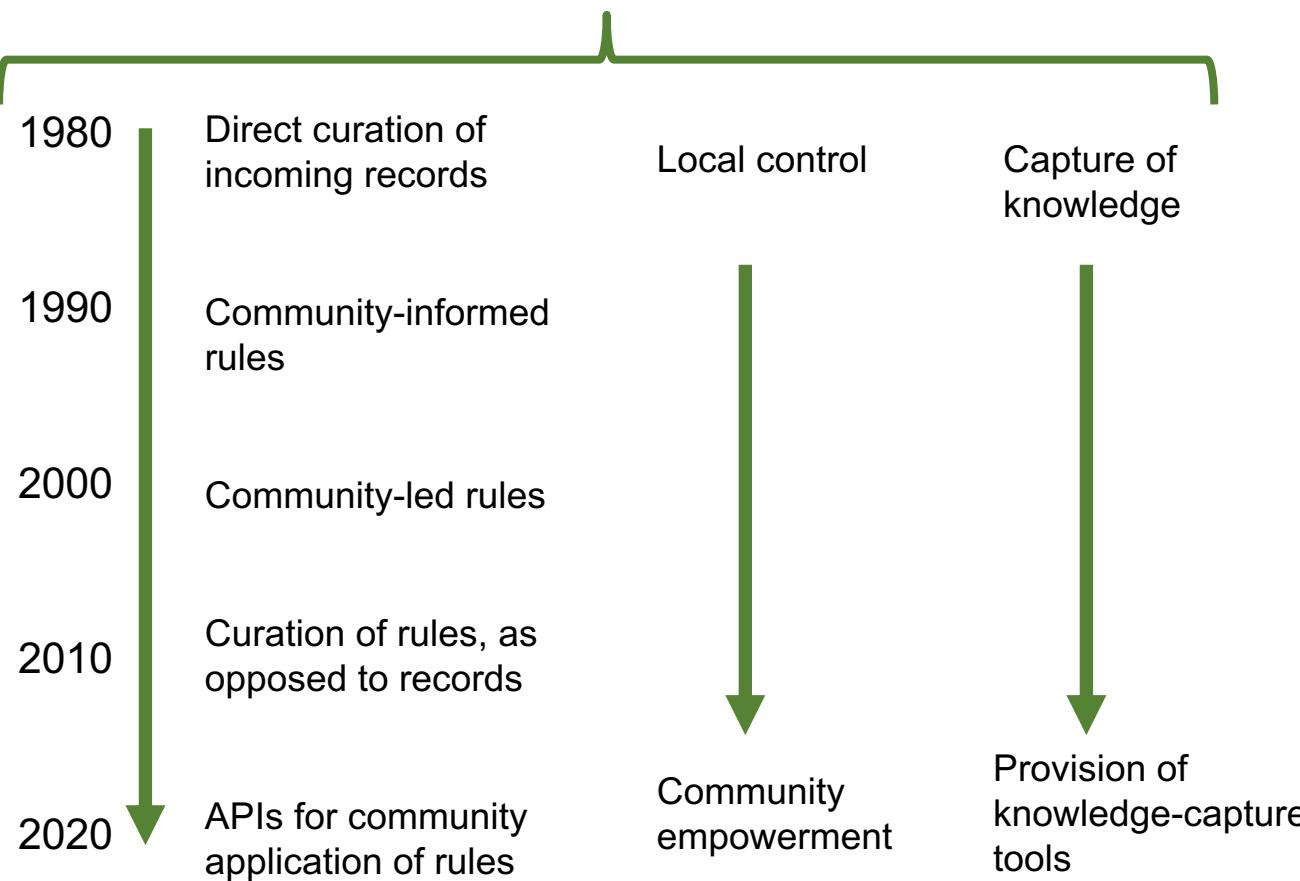
EMBL European Bioinformatics Institute



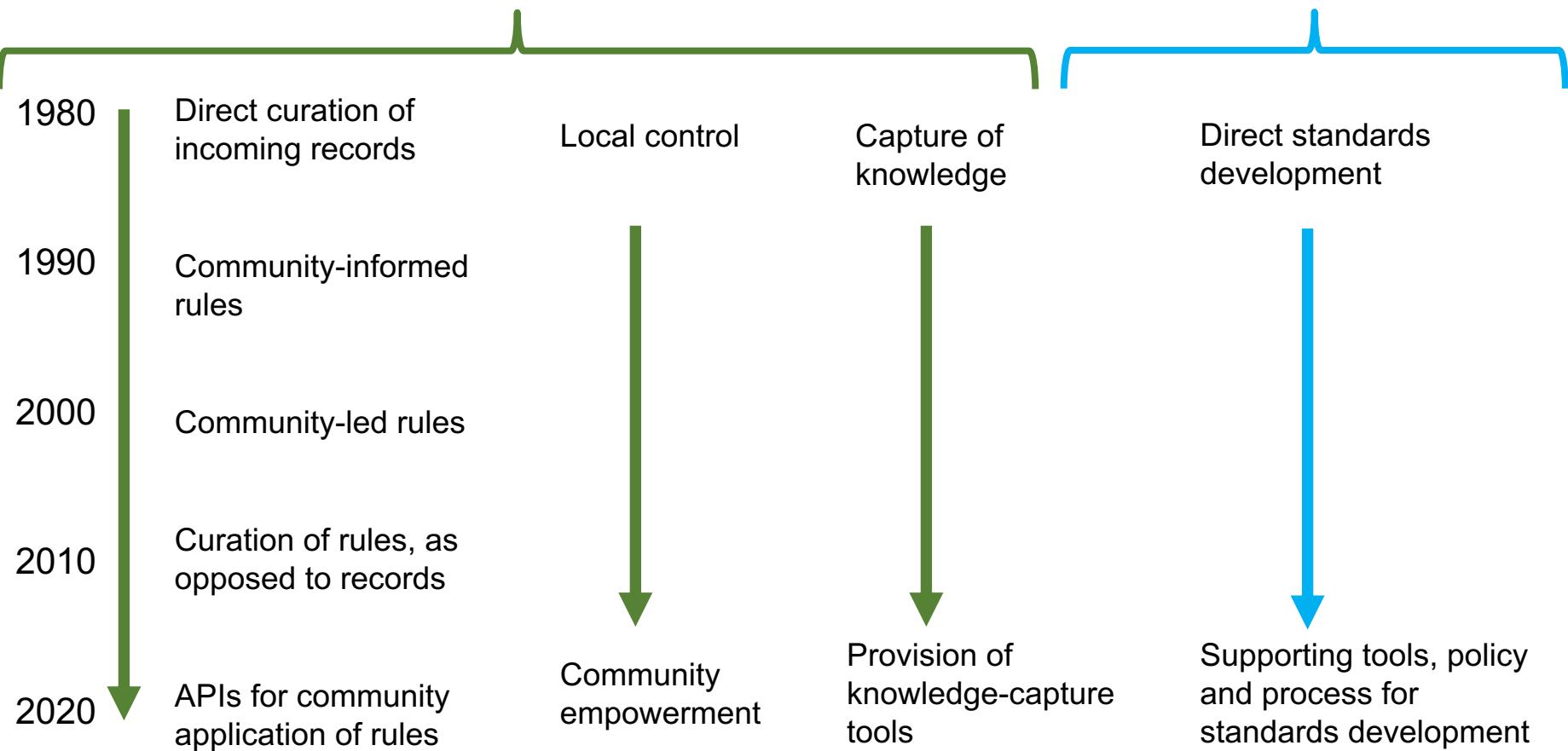
Nucleotide sequencing scale



Response to growth



Response to growth



Metagenome Exchange registry

MG-RAST metagenomics analysis server

The oral metagenome in health and disease (mpg128)

principle investigator: Alex Mira, CSISP
 visibility: public
 static link: [http://metagenomics.anl.gov/linkin.cgi?project=mpg128](http://metagenomics.anl.gov/mgmain.html?mgpage=project&project=mpg128)
 ENA link: [ERP021171](#)
 description:
 The oral cavity of humans is inhabited by hundreds of bacterial species and some of them have a key role in the development of oral diseases, mainly dental caries and periodontitis. We describe for the first time the metagenome of the human oral cavity under health and diseased conditions, with a focus on supragingival dental plaque and cavities. Direct pyrosequencing of eight samples with different oral-health status produced 1 Gbp of sequence without the biases imposed by PCR or cloning. These data show that cavities are not dominated by St species originally identified as the etiological agent of dental caries) but are in fact a complex community formed by tens of bacterial species, in agreement with the polymicrobial disease. The analysis of the reads indicated that the oral cavity is functionally a different environment from the gut, with many functional categories enriched in one of the two environments and depicted in the other. Individuals who had never suffered from dental caries showed an over-representation of several functional categories, like genes for antimicrobial peptides and quorum sensing. In addition, they did not have many of other species. Several isolates belonging to these dominant bacteria in healthy individuals were cultured and shown to inhibit the growth of cariogenic bacteria, suggesting the use of these commensal bacterial strains as probiotics to promote oral health and prevent dental caries.

funding source:
 Spanish MICINN: SAF2009-13032-C02-02 from the I+D program, BIO2008-CMICROGEN CSD2009-00006 from the Consolider- Ingenio program.

contact:
 Administrative
 Alex Mira (mira_ale@gva.es)
 CSISP (<http://www.csisp.gva.es/web/alex/home>)
 Avda. Cataluña, 21. Valencia, Spain

Technical
 Pedro Belida-Ferre (belida_ped@gva.es)
 Center for Advanced Research in Public Health, Department of Genomics and
 Avda. Cataluña, 21 ; 46020 ; Valencia ; Comunidad Valenciana, Spain

[metagenomes](#)

<https://www.ebi.ac.uk/ena/data/view/PRJEB19188>

Study: PRJEB19188
 The oral metagenome in health and disease
 View: Project XML, Study XML

Name: mpg128
 Secondary accession(s): ERP021171
 Broker Name: MG-RAST
 Description:
 The oral cavity of humans is inhabited by hundreds of bacterial species and some of them have a key role in the development of oral diseases, mainly dental caries and periodontitis. We describe for the first time the metagenome of the human oral cavity under health and diseased conditions, with a focus on supragingival dental plaque and cavities. Direct pyrosequencing of eight samples with different oral-health status produced 1 Gbp of sequence without the biases imposed by PCR or cloning. These data show that cavities are not dominated by St species originally identified as the etiological agent of dental caries) but are in fact a complex community formed by tens of bacterial species, in agreement with the polymicrobial disease. The analysis of the reads indicated that the oral cavity is functionally a different environment from the gut, with many functional categories enriched in one of the two environments and depicted in the other. Individuals who had never suffered from dental caries showed an over-representation of several functional categories, like genes for antimicrobial peptides and quorum sensing. In addition, they did not have many of other species. Several isolates belonging to these dominant bacteria in healthy individuals were cultured and shown to inhibit the growth of cariogenic bacteria, suggesting the use of these commensal bacterial strains as probiotics to promote oral health and prevent dental caries.

Navigation Read Files Portal Attributes

Bulk Download Files (If the app doesn't open, please use Firefox to launch the bulk downloader app.)

Showing results 1 - 8 of 8 results

Study accession	Sample accession	Secondary sample accession	Experiment accession	Run accession	Tax ID	Scientific name	Instrument model	Library layout	FASTQ files (FTP)	FASTQ files (Galaxy)	Submitted files (FTP)	Submitted files (Galaxy)	NCBI SRA file (FTP)	NCBI SRA file (Galaxy)	CRAM Index files (FTP)	CRAM Index files (Galaxy)
PRJEB19188	SAMEA54122668	ERS1510278	ERX1806527	ERR1806500	447426	human oral metagenome	unspecified	SINGLE	Fasta File 1	Fasta File 1						
PRJEB19188	SAMEA54121168	ERS1510278	ERX1806528	ERR1806501	447426	human oral metagenome	unspecified	SINGLE	Fasta File 1	Fasta File 1						
PRJEB19188	SAMEA54125668	ERS1510282	ERX1806529	ERR1806502	447426	human oral metagenome	unspecified	SINGLE	Fasta File 1	Fasta File 1						
PRJEB19188	SAMEA54126418	ERS1510283	ERX1806530	ERR1806503	447426	human oral metagenome	unspecified	SINGLE	Fasta File 1	Fasta File 1						
PRJEB19188	SAMEA54124918	ERS1510281	ERX1806531	ERR1806504	447426	human oral metagenome	unspecified	SINGLE	Fasta File 1	Fasta File 1						

<https://www.ebi.ac.uk/metagenomics/projects/ERP021171>; <https://www.mg-rast.org/mgmain.html?mgpage=project&project=mpg128>; <https://www.ebi.ac.uk/ena/data/view/PRJEB19188>

Project: The oral metagenome

EBI Metagenomics

Home Search Sequence search Submit data Projects Samples Comparison tool About Help

Not logged in Login

Project ERP021171 (PRJEB19188)
 The oral metagenome in health and disease

Overview Analysis summary

Last updated: 14-Feb-2017

Description

The oral cavity of humans is inhabited by hundreds of bacterial species and some of them have a key role in the development of oral diseases, mainly dental caries and periodontitis. We describe for the first time the metagenome of the human oral cavity under health and diseased conditions, with a focus on supragingival dental plaque and cavities. Direct pyrosequencing of eight samples with different oral-health status produced 1 Gbp of sequence without the biases imposed by PCR or cloning. These data show that cavities are not dominated by Streptococcus mutans (the species originally identified as the etiological agent of dental caries) but are in fact a complex community formed by tens of bacterial species, in agreement with the polymicrobial disease. The analysis of the reads indicated that the oral cavity is functionally a different environment from the gut, with many functional categories enriched in one of the two environments and depicted in the other. Individuals who had never suffered from dental caries showed an over-representation of several functional categories, like genes for antimicrobial peptides and quorum sensing. In addition, they did not have mutans streptococci but displayed high recruitment of other species. Several isolates belonging to these dominant bacteria in healthy individuals were cultured and shown to inhibit the growth of cariogenic bacteria, suggesting the use of these commensal bacterial strains as probiotics to promote oral health and prevent dental caries.

Classification: Host-associated > Human > Digestive system > Oral

Submitter details

Institute: CSISP

Associated runs

Show 25 50 results Download detailed info (CSV) Filter:

Sample Name	Sample ID	Run ID	Experiment type	Instrument model	Version	Analysis results
mgas2920 Taxonomy ID:447426	ERS1510276	ERR1806501	Metagenomic	unspecified	3.0	Taxonomy Function ↓
mgas2922 Taxonomy ID:447426	ERS1510277	ERR1806507	Metagenomic	unspecified	3.0	Taxonomy Function ↓
mgas2923 Taxonomy ID:447426	ERS1510278	ERR1806500	Metagenomic	unspecified	3.0	Taxonomy Function ↓
mgas2924 Taxonomy ID:447426	ERS1510279	ERR1806508	Metagenomic	unspecified	3.0	Taxonomy Function ↓
mgas2925 Taxonomy ID:447426	ERS1510280	ERR1806505	Metagenomic	unspecified	3.0	Taxonomy Function ↓
mgas2926 Taxonomy ID:447426	ERS1510281	ERR1806504	Metagenomic	unspecified	3.0	Taxonomy Function ↓
mgas2927 Taxonomy ID:447426	ERS1510282	ERR1806502	Metagenomic	unspecified	3.0	Taxonomy Function ↓
mgas2928 Taxonomy ID:447426	ERS1510283	ERR1806503	Metagenomic	unspecified	3.0	Taxonomy Function ↓

Showing 1 to 8 of 8 results

Common data sets

- Analysis based on PI name, centre name and project name

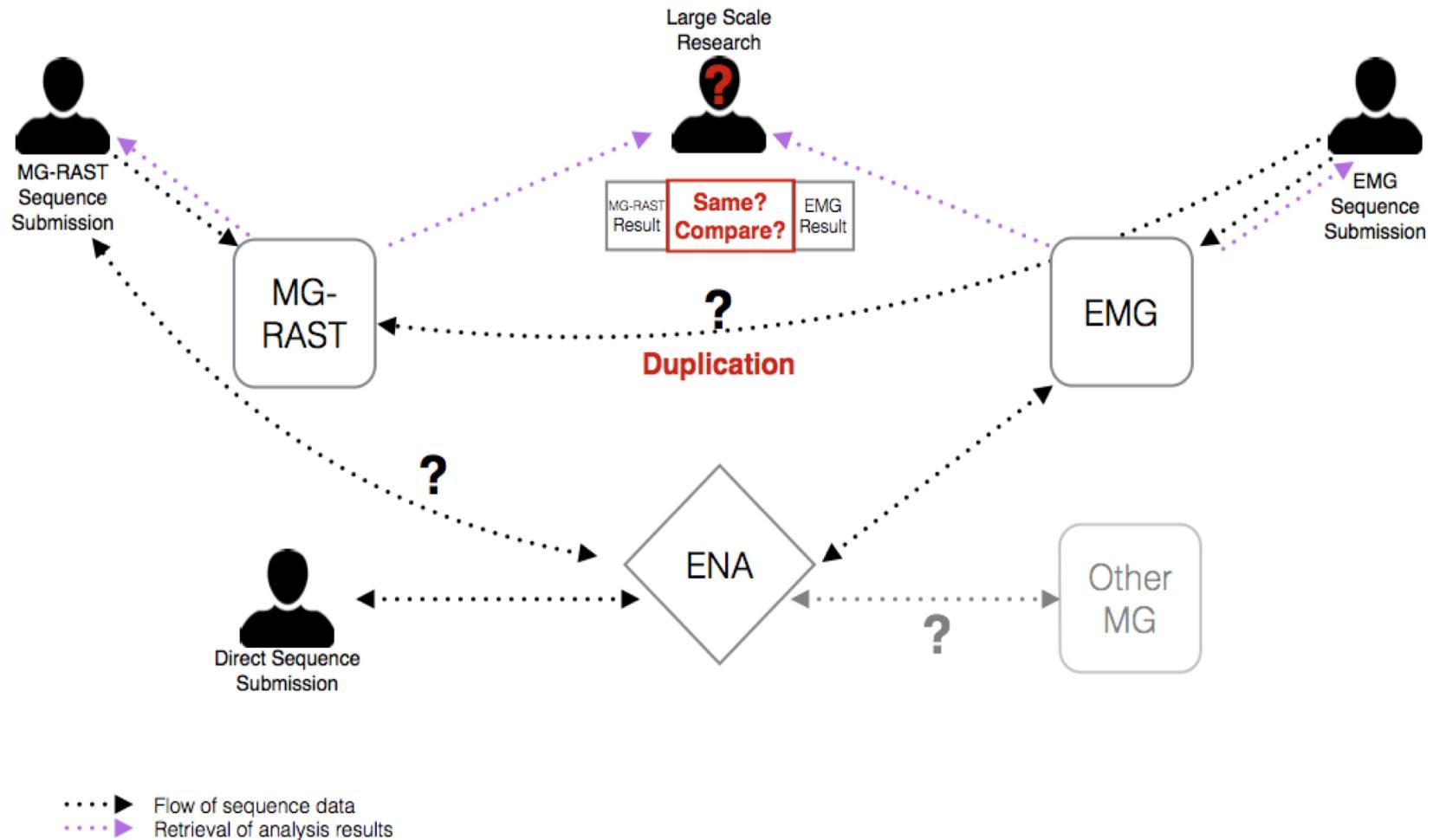
same_PI: Francisco Massot				
MG_RAST	Hasselt University		mpg81386	Metagenomic studies on pasturelands chronically exposed to pesticides
ENA	Centre for Environmental Sciences, Hasselt University		PRJEB22053	Metagenomic study on pasturelands chronically exposed to pesticides

same_PI: Carolina Reyes				
MG_RAST	University of Vienna		mpg10811	Baltic Sea Project Metagenome
ENA	UNIVERSITY OF VIENNA		PRJEB9348	Pyrosequencing study of the Skagerrak and Bothnian Bay

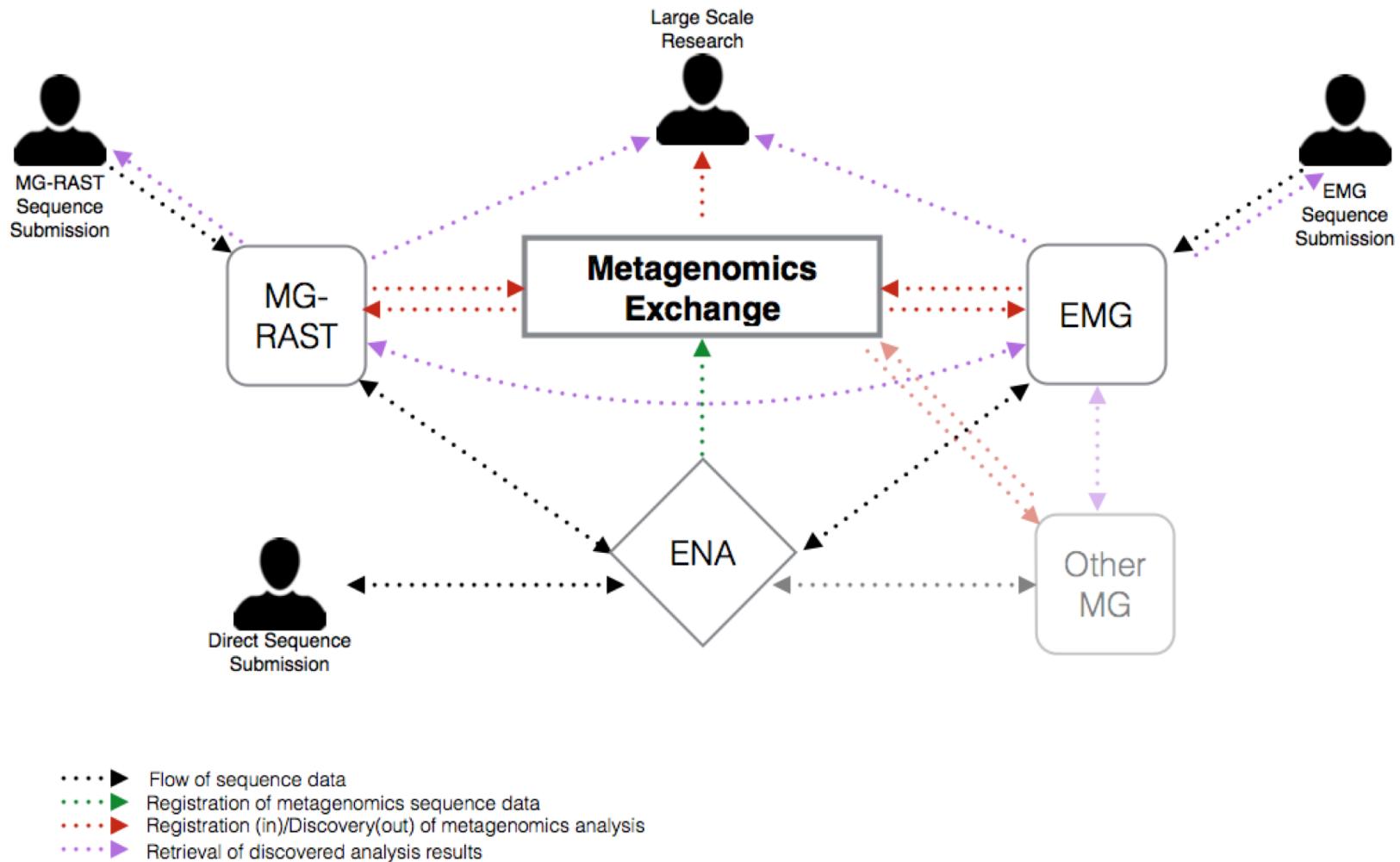
same_PI: Bonnie Brown				
MG_RAST	VCU Biology		mpg18456	MinION longreads for diagnosis of human ailments
MG_RAST	VCU Biology		mpg20278	MinION Global River Sequencing
MG_RAST	Virginia Commonwealth University		mpg12307	Experimental Metagenome on MinION
MG_RAST	Virginia Commonwealth University		mpg5507	James River Epidemiological Metagenome
ENA	Virginia Commonwealth University		PRJEB14105	?29 Treatment for MinION
ENA	Virginia Commonwealth University		PRJEB15077	Global Rivers
ENA	Virginia Commonwealth University		PRJEB8672	Experimental Metagenome on MinION
ENA	Virginia Commonwealth University		PRJEB8716	Experimental Metagenome on MinION

- Overall: identified 120 MG-RAST projects link to 321 ENA projects
- Scale of EMG and MG-RAST in hundreds of thousands

Metagenomics Exchange



Metagenomics Exchange

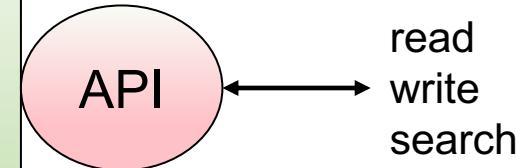


- Flow of sequence data
- Registration of metagenomics sequence data
- Registration (in)/Discovery(out) of metagenomics analysis
- Retrieval of discovered analysis results

Metagenomics Exchange Registry



	MGX ID (MGX#####) RUN ID (ERR#####) RUN status (public/private/suppressed) data set/analysis end point broker (EMG/ENA/MG-RAST) user ID for data set data set endpoint method confidence created date (in registry) <i>modified date (in registry)</i>
--	--



- ENA writes entry for new data sets
- ME partners discover data sets
- ME partners report endpoints for analyses of data sets
- Extensions beyond EMBL-EBI Metagenomics and MG-RAST
- Imminent launch

Contextual data clearinghouse



Home | Search & Browse | Submit & Update | Software | About ENA | Support

Sample: SAMN00017459

Generic sample from *Archaeoglobus profundus* DSM 5631

View: XML

Name 796	Submitting Centre DOE Joint Genome Institute
Secondary accession(s) SRS085412	
Broker Name NCBI	
Description <i>Archaeoglobus profundus</i> DSM 5631	
Lineage Archaea, Euryarchaeota, Archaeoglobi, Archaeoglobales, Archaeoglobaceae, Archaeoglobus	
Navigation	Read Files
Attributes	
Strain DSM 5631	
Sample Number 4085236	

Archaeoglobus profundus DSM 5631

[+Expand all](#) / [-Collapse all](#)

- Summary

MMP ID	MMP00017459
Full Scientific Name	<i>Archaeoglobus profundus</i> DSM 5631
Strain	Av18, DSM 5631, JCM 9629, NBRC 100127
Type Strain	Yes
Geographic location	Mexico, Gulf of California, Guaymas Basin
Collection Date	<i>missing</i>
ENA BioSample accession identifier	SAMN00017459
Bacdive ID	18097
Culture Collection(s)	DSM 5631, JCM 9629, NBRC 100127
Isolation Country	Mexico
Environmental Package	Sediment
Isolation Source	Deep-sea hydrothermal vent
Host Scientific Name	<i>not applicable</i>
Curation Date	2016-05-20
Updated Date	<i>missing</i>
Implementation Date	2016-11-22



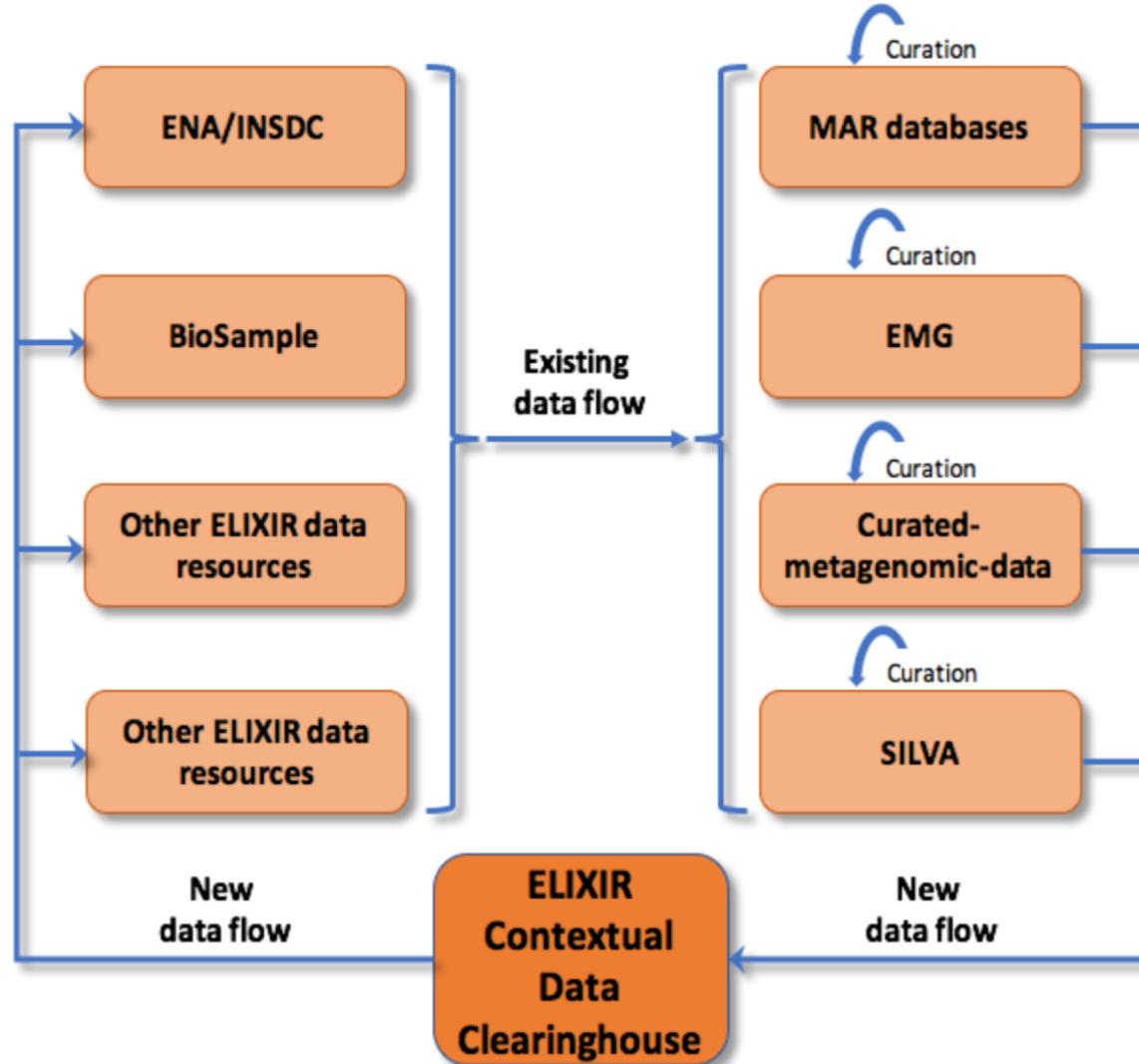
MARREF

Contextual data clearinghouse



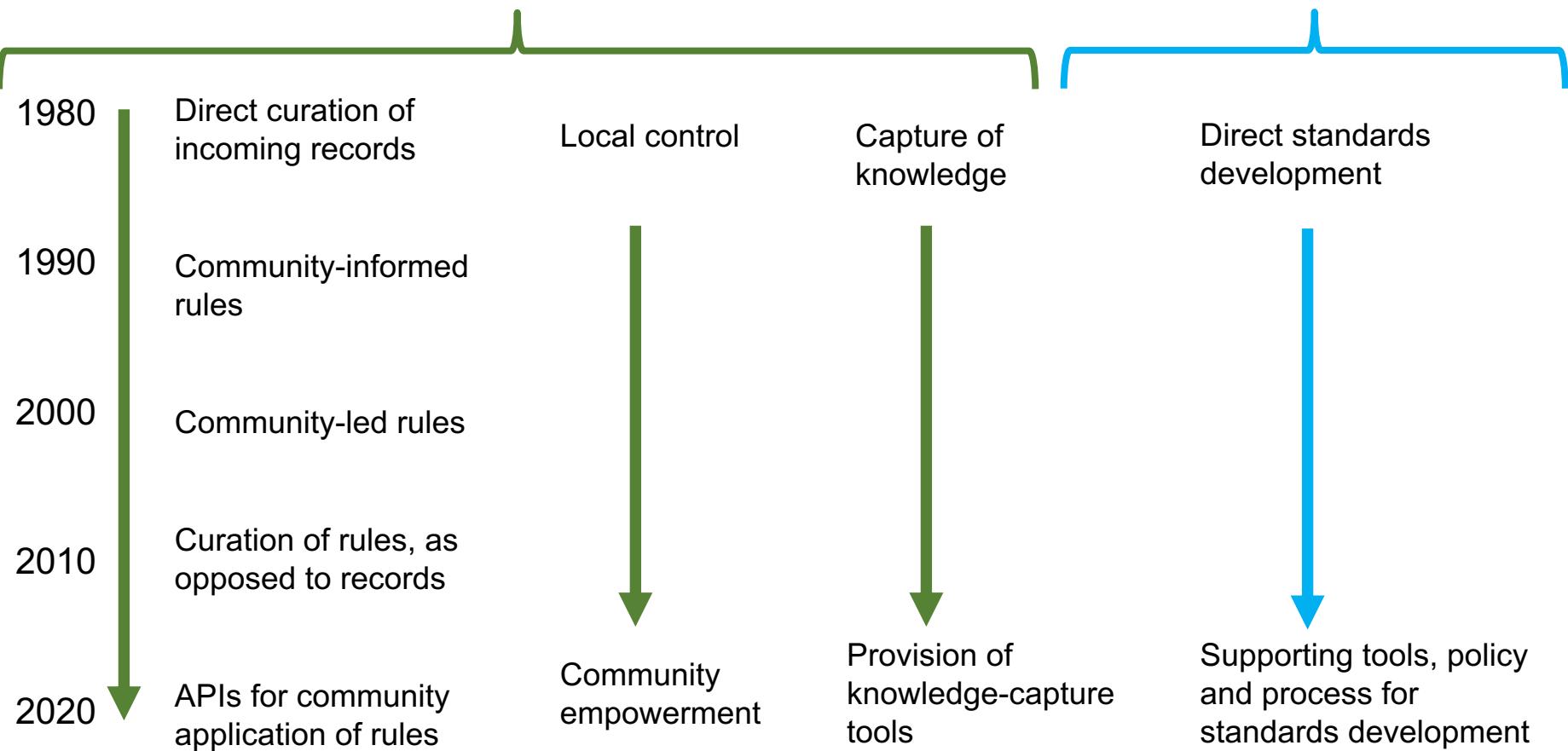
- Many groups operate processes that draw data from archives and add include contextual and other metadata curation
 - E.G. MAR databases, SILVA, EMBL-EBI Metagenomics, curatedMetagenomicData, etc.
 - Archival workflows for addition/update/correction are slow
- This work is valuable, but often redundant and of limited impact
- Decoupling reporting from archival updates will enable

Decoupled reporting and update



- API for reporting
- API for search/retrieval
- Focus on workflows at BioSamples, ENA
- Open to other providers and consumers

Response to growth



Acknowledgements

- **Content & support:** Ana Cerdeño-Tárraga, Ana Luisa Toribio, Annelisa Milano, Petra ten Hoopen, Marc Rosello, Jeena Rajan, Isabel Santos Magalhaes, **Clara Amid**; Ian Streeter, Susan Fairley, David Richardson, **Peter Harrison**, **Laura Clarke**
- **Technology (presentation & collaborative tools):** Blaise Alako, Simon Kay, **Nima Pakseresht**; Xin Liu, Suran Jayathilaka, **Nicole Silvester**
- **Technology (submissions & data back-end):** Daniel Vaughan, Neil Goodgame, Iain Cleland, Josué Martinez Villacorta, Dmitriy Smirnov, Kethi Reddy, Vadim Zalunin, *James Bonfield*, Rasko Leinonen, **Thomas Keane**
- **Infrastructure:** EMBL-EBI Technical Services Cluster
- **Metagenome Exchange (ANL):** William Trimble, Andreas Wilke, **Folker Meyer**
- **Metagenome Exchange (EMBL-EBI):** Alex Mitchell, Ola Tarkowska, **Rob Finn**
- **Contextual data clearinghouse:** Nils Peder Willlassen (UiT), Nicola Segata (Uni. Bari), Graziano Pesole (Uni. Bari), Pelin Yilmaz (MPI-Bremen), Frank Oliver Glöckner (MPI-Bremen), Helen Parkinson (EMBL-EBI), Robb Finn (EMBL-EBI), Tony Burdett (EMBL-EBI)
- **INSDC Partners:** NCBI, DDBJ
- **Funders**

