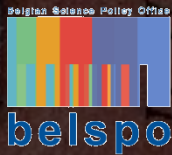


ICCC12 Conference 2010

Costão do Santinho International Events Center, Florianópolis, Santa Catarina, Brazil, 26 September-1 October 2010

S T R A I N I N F O . N E T

providing passports to cultured organisms







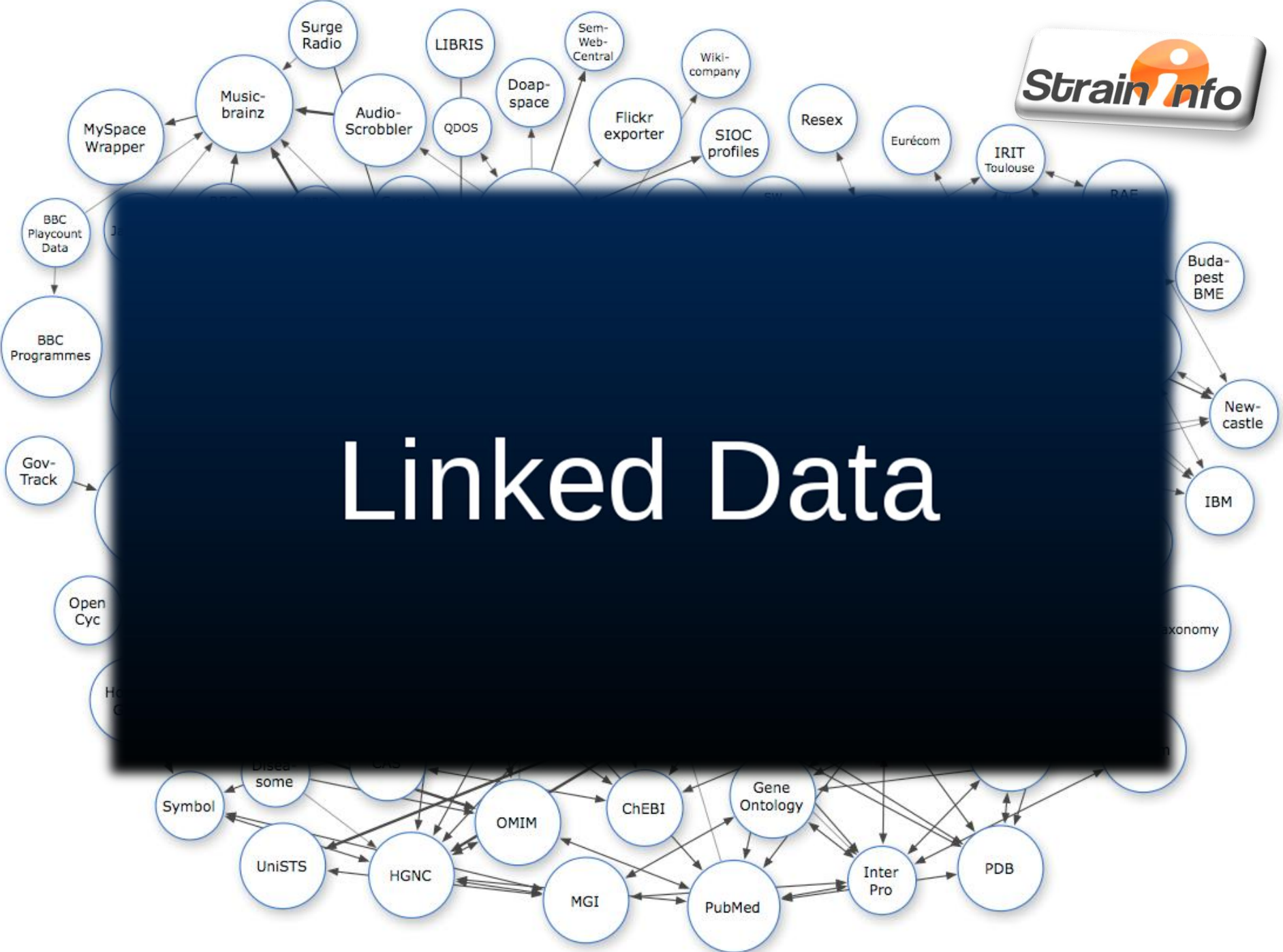


Linked Data

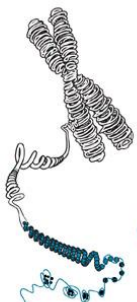
WWW@20

W3C

13 March 2009



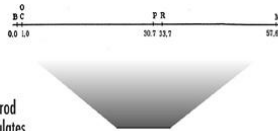
Linked Data



Mendel discovers laws of genetics
1865



Rediscovery of Mendel's work
1900



Garrad formulates the concept of human inborn errors of metabolism
1905

Sturtevant makes the first linear map of genes
1913

Avery, McLeod, and McCarty demonstrate DNA is the hereditary material
1944



Watson and Crick describe the double helical structure of DNA
1953

	U	C	A	G
Phe	Leu	Ser	Lys	Cys
Leu	Pro	His	Glu	Arg
Ile	Met	Thr	Asn	Ser
G	Val	Ala	Asp	Gly



Cohen and Boyer develop recombinant DNA technology
1972

Issuing of Belmont Report on the use of human subjects in research
1974

Sanger and Maxam & Gilbert develop DNA sequencing methods
1977

GenBank database established
1982



First human gene with DNA -- Huntington's
1993

1990

1991

1992

1993

1994

1995

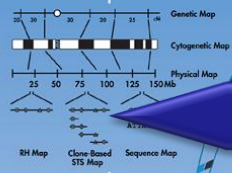
1996

1997

1998

Human Genome Project (HGP) launched in the U.S.

First U.S. Genome Centers established



Second-generation human genetic

1995

Craig Venter
TIGR



complete genome of bacteria
Haemophilus influenzae

HGP's human physical mapping goal achieved

First human gene map established

DOE forms Joint Genome Institute

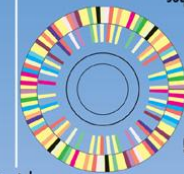
Incorporation of genes into human genome



NCHGR becomes NHGRI



Pilot projects for human genome sequencing begin in U.S.



First bacterial genome (*Haemophilus influenzae*) sequenced

First archaeal genome sequenced

Yeast (*S. cerevisiae*) genome sequenced

E. coli genome sequenced

New five-year plan for the HGP in the U.S.



RIKEN Genome Center (Japan)

Roundworm (*C. elegans*) genome sequenced

Ethical, Legal, and Social Implications (ELSI) programs founded at NIH and DOE

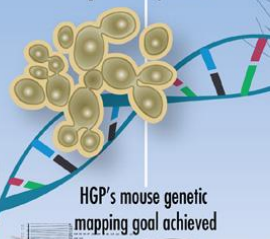
Rapid data release guidelines established by NIH and DOE

Sanger Centre founded (later renamed Wellcome Trust Sanger Institute)



The Wellcome Trust

U.S. Equal Employment Opportunity Commission issues policy on genetic discrimination in the workplace



HGP's mouse genetic mapping goal achieved

Genoscope (French National Genome Sequencing Center) founded

SNP initiative

GTG
GTC

First gene for breast cancer (BRCA1) mapped



Bermuda principles for rapid and open data release established

Chinese National Human Genome Center (in Beijing and Shanghai)

Collins F.S. et al. (2003). *Nature* 422, 835-847.

H. influenzae genome



Whole-Genome Random Sequencing and Assembly of *Haemophilus influenzae* Rd

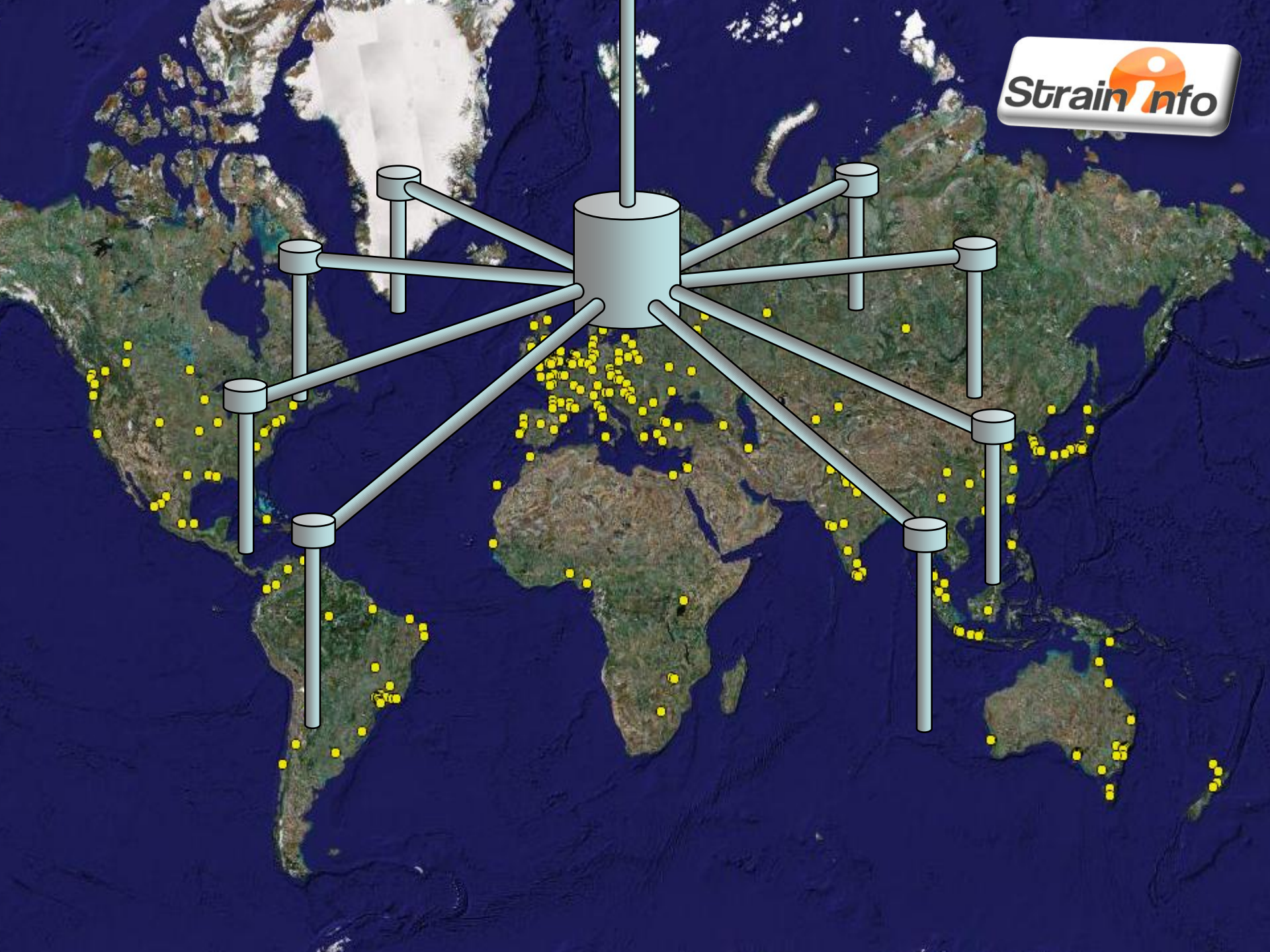
Robert D. Fleischmann, Mark D. Adams, Owen White, Rebecca A. Clayton, Ewen F. Kirkness, Anthony R. Kerlavage, Carol J. Bult, Jean-Francois Tomb, Brian A. Dougherty, Joseph M. Merrick, Keith McKenney, Granger Sutton, Will FitzHugh, Chris Fields,* Jeannine D. Gocayne, John Scott, Robert Shirley, Li-Ing Liu, Anna Glodek, Jenny M. Kelley, Janice F. Weidman, Cheryl A. Phillips, Tracy Spriggs, Eva Hedblom, Matthew D. Cotton, Teresa R. Utterback, Michael C. Hanna, David T. Nguyen, Deborah M. Saudek, Rhonda C. Brandon, Leah D. Fine, Janice L. Fritchman, Joyce L. Fuhrmann, N. S. M. Geoghagen, Cheryl L. Gnehm, Lisa A. McDonald, Keith V. Small, Claire M. Fraser, Hamilton O. Smith, J. Craig Ventert†

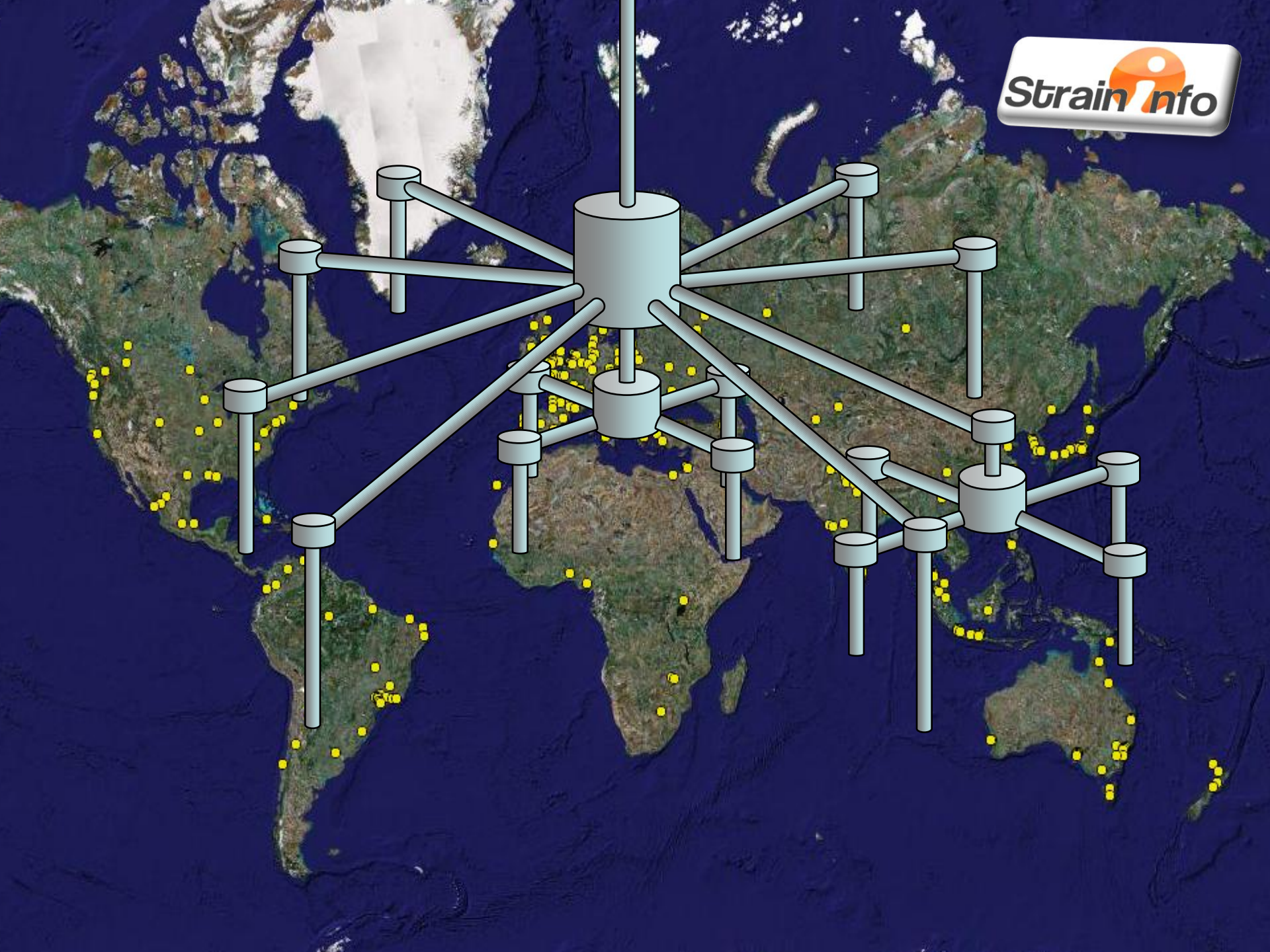
SCIENCE • VOL. 269 • 28 JULY 1995

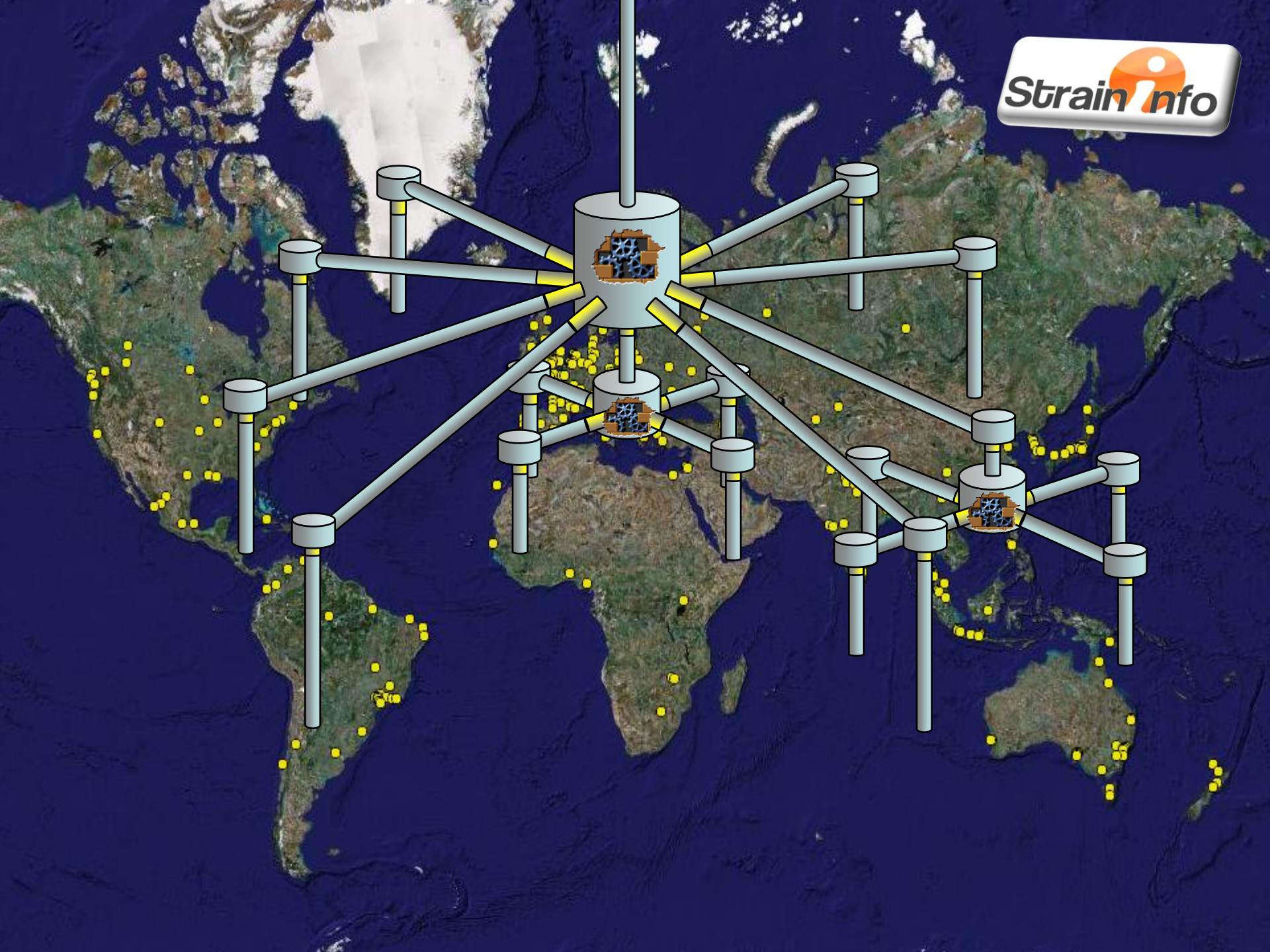
Haemophilus influenzae Rd genome was the first to be sequenced and assembled in a free-living organism. It contains about 1.8 million base pairs and is estimated to have 1,740 genes.

Haemophilus influenzae Rd is a non-motile Gram-negative bacterium. It was first identified in 1892 by Dr. Richard Griffith. It is a facultative anaerobe with a wide range of host specificity.









H. influenzae genome

A screenshot of a web browser window displaying a presentation slide. The browser's address bar shows the URL <http://www.straininfo.net/>. The slide content includes the text "Rd" in the top left, "Linked Data" in large white font in the center, and a footer with logos for "KERMIT" and "UNIVERSITEIT GENT".

Rd

Linked Data

KERMIT UNIVERSITEIT GENT



H. influenzae genome



Whole-Genome Random

S

Search

al. 1917 -

Linked Data

496

CONSULTANTS

41111 83111 11111

No 16S rDNA sequence in SQX for CCUG 18800

H. influenzae genome



Linked Data

NCBI Resources How To My NCBI Sign In

Nucleotide Alphabet of Life

Search: Nucleotide Limits Advanced search Help

Search Clear

Send Change region shown

Display Settings

Sequence

Haemophilus influenzae

GenBank: L42023

FASTA

Comment

LOCUS

DEFINITION

ACCESSION

VERSION

DBLINK

KEYWORDS

SOURCE

ORGANISM

REFERENCE

AUTHOR

TITLE

JOURNAL

PUBLISHED

Print this Page

Cross References:

Nucleotide (GenBank) : [AJ277814](#) Haemophilus influenzae waaA gene for KDO transferase, strain DSM 11121.

Nucleotide (GenBank) : [AJ277816](#) Haemophilus influenzae kdsA gene for KDO kinase, strain 11121.

Nucleotide (GenBank) : [L42023](#) Haemophilus influenzae Rd KW20.

BioServices
Bio-materials management;
basic repository to complex
partnership-level services

Automatic 16S rRNA tree builder



Ensifer sp. Taxon Pas... x

http://www.straininfo.net/taxa/3561

StrainInfo BETA [news](#) [about](#) [contact](#) [help](#)

log in | register

Advanced search StrainInfo Projects

Taxon Passport

Ensifer sp.

overview

genus *Ensifer* sp.
external links [DSMZ](#), [LPSN](#); [J.P. Euzéby](#), [NCBI](#)
search StrainInfo [Find all strains](#) [Find all type strains](#)

subtaxa

12 items found, displaying all items.

- Ensifer adhaerens*
- Ensifer arboris*
- Ensifer fredii*
- Ensifer garamanticus*
- Ensifer kostiensis*
- Ensifer kummerowiae*
- Ensifer medicae*
- Ensifer melliloti*
- Ensifer numidicus*
- Ensifer saheli*
- Ensifer terangae*
- Ensifer xinjiangensis*

12 items found, displaying all items.

phylogenetic tree (16S)

```
graph TD
    Root --- S_morelense["S. morelense LMG 21331†"]
    Root --- E_sp["E. sp. LMG 20571"]
    Root --- E_adhaerens_A["E. adhaerens gv. A LMG 9954"]
    Root --- E_adhaerens_C["E. adhaerens gv. C LMG 20216†"]
    Root --- E_adhaerens_B["E. adhaerens gv. B R-7457"]
    Root --- E_terangae["E. terangae LMG 7834†"]
    Root --- E_fredii["E. fredii LMG 6217†"]
    Root --- S_americanum["S. americanum LMG 22684†"]
    Root --- E_saheli["E. saheli LMG 7837†"]
    Root --- E_medicae["E. medicae LMG 19920†"]
    Root --- E_arboris["E. arboris LMG 14919†"]
    Root --- E_melliloti["E. melliloti LMG 6133†"]
    Root --- E_kostiensis["E. kostiensis LMG 19227†"]
    Root --- C_crescentia["C. crescentia CB15"]
```

publications

Young JM
The genus name *Ensifer* Casida 1982 takes priority over *Sinorhizobium* Chen et al. 1988, and *Sinorhizobium morelense* Wang et al. 2002 is a later synonym of *Ensifer adhaerens* Casida 1982. Is the combination "*Sinorhizobium adhaerens*" (Casida 1982) Willems et al. 2003 legitimate? Request for an Opinion

Automatic 16S rRNA tree builder



- check latest taxonomic status of genus
- select 16S sequences for each type strain
- download 16S sequences
- create multiFASTA file
- compute multiple alignment
- compute genetic distances
- compute and render phylogenetic tree

Automatic 16S rRNA tree builder

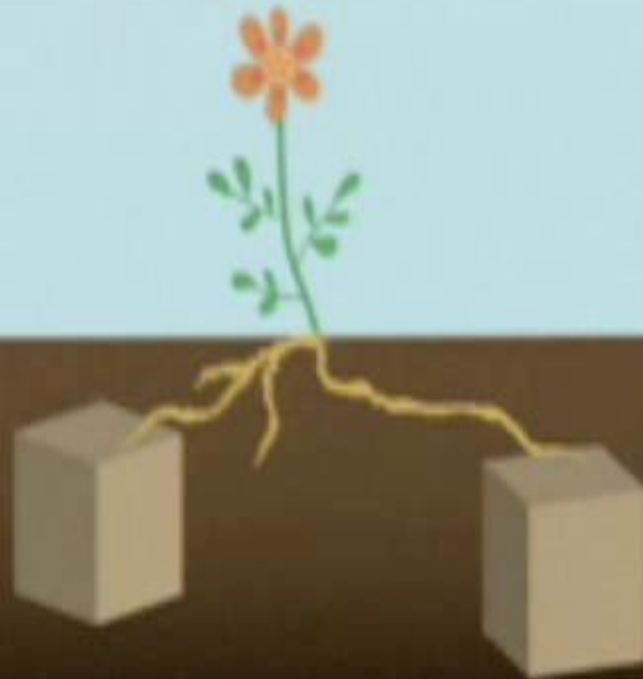


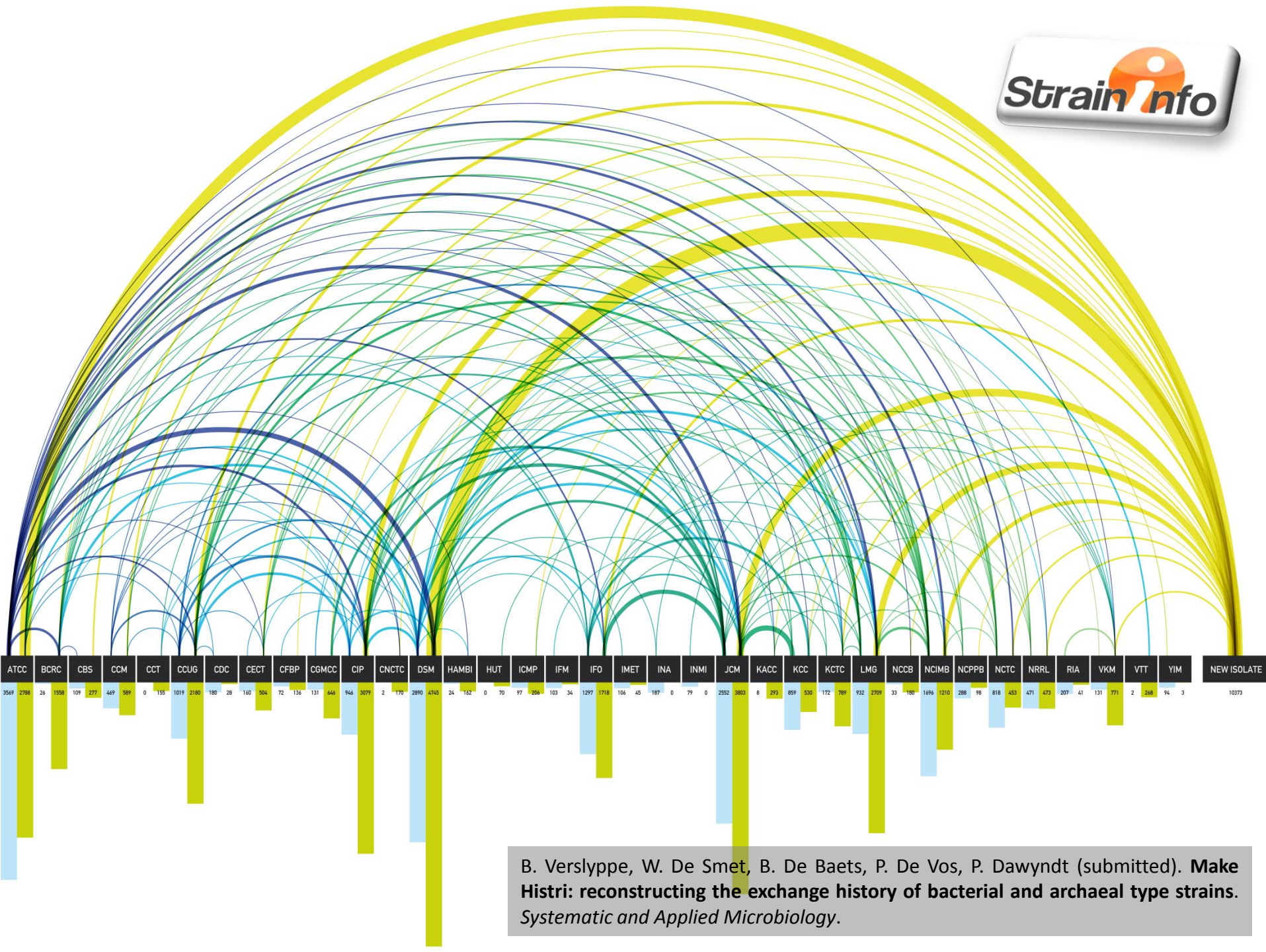
input

Linked Data

phylogenetic
tree

demo





B. Verslyppe, W. De Smet, B. De Baets, P. De Vos, P. Dawyndt (submitted). **Make Histri: reconstructing the exchange history of bacterial and archaeal type strains.** *Systematic and Applied Microbiology*.



LINKED DATA STANDARDS



Microbiological Common Language



INSTITUT PASTEUR

Research in Microbiology 161 (2010) 439–445

www.elsevier.com/locate/resmic

Microbiological Common Language (MCL): a standard for electronic information exchange in the Microbial Commons

Bert Verslyppe^{a,b}, Renzo Kottmann^c, Wim De Smet^a, Bernard De Baets^d,
Paul De Vos^{a,e}, Peter Dawyndt^{b,*}

Verslyppe B., Kottman R., De Vos P., De Baets B. & Dawyndt P. (2010). **MCL: a microbiological language for electronic information exchange in the Microbial Commons.** *Research in Microbiology* 161, 439-445.

Microbiological Common Language



```
<si:Culture>
  <si:strainNumber>LMG 24056</si:strainNumber>
  <si:otherStrainNumber>DSM 44871</si:otherStrainNumber>
  <si:otherStrainNumber>Trujillo LUPAC 09</si:otherStrainNumber>
  <si:catalogURL>http://bccm.belspo.be/db/lmg_strain_details.php?NUM=24056</si:catalogURL>
  <si:cultureLastUpdateDate>2008-08-05T12:30:00</si:cultureLastUpdateDate>

  <si:speciesName>Micromonospora saelicesensis</si:speciesName>
  <si:nomenclaturalPublication>
    <dcterms:bibliographicCitation>Trujillo, Kroppenstedt, Fernandez-Moliner, Schumann and Martinez-Molina 2007</dcterms:bibliographicCitation>
  </si:nomenclaturalPublication>

  <si:isolationDate>2003</si:isolationDate>
  <si:isolator>M.Trujillo</si:isolator>
  <si:isolatorInstitute>Dep. de Microbiologia y Genetica Universidad de Salamanca</si:isolatorInstitute>
  <si:Sample>
    <si:sampleLocationCountry>Spain</si:sampleLocationCountry>
    <si:sampleLocationPlace>Salamanca</si:sampleLocationPlace>
    <si:sampleHabitat>Lupinus angustifolius, root nodule</si:sampleHabitat>
  </si:Sample>

  <si:Deposit>
    <si:resultingStrainNumber>LMG 24056</si:resultingStrainNumber>
    <si:depositDate>2007</si:depositDate>
    <si:depositor>M.Trujillo</si:depositor>
    <si:depositorInstitute>Dep. de Microbiologia y Genetica Universidad de Salamanca</si:depositorInstitute>
  </si:Deposit>
  <si:history>&lt;- 2007, M.Trujillo Dep. de Microbiologia y Genetica Universidad de Salamanca Spain (2003)</si:history>

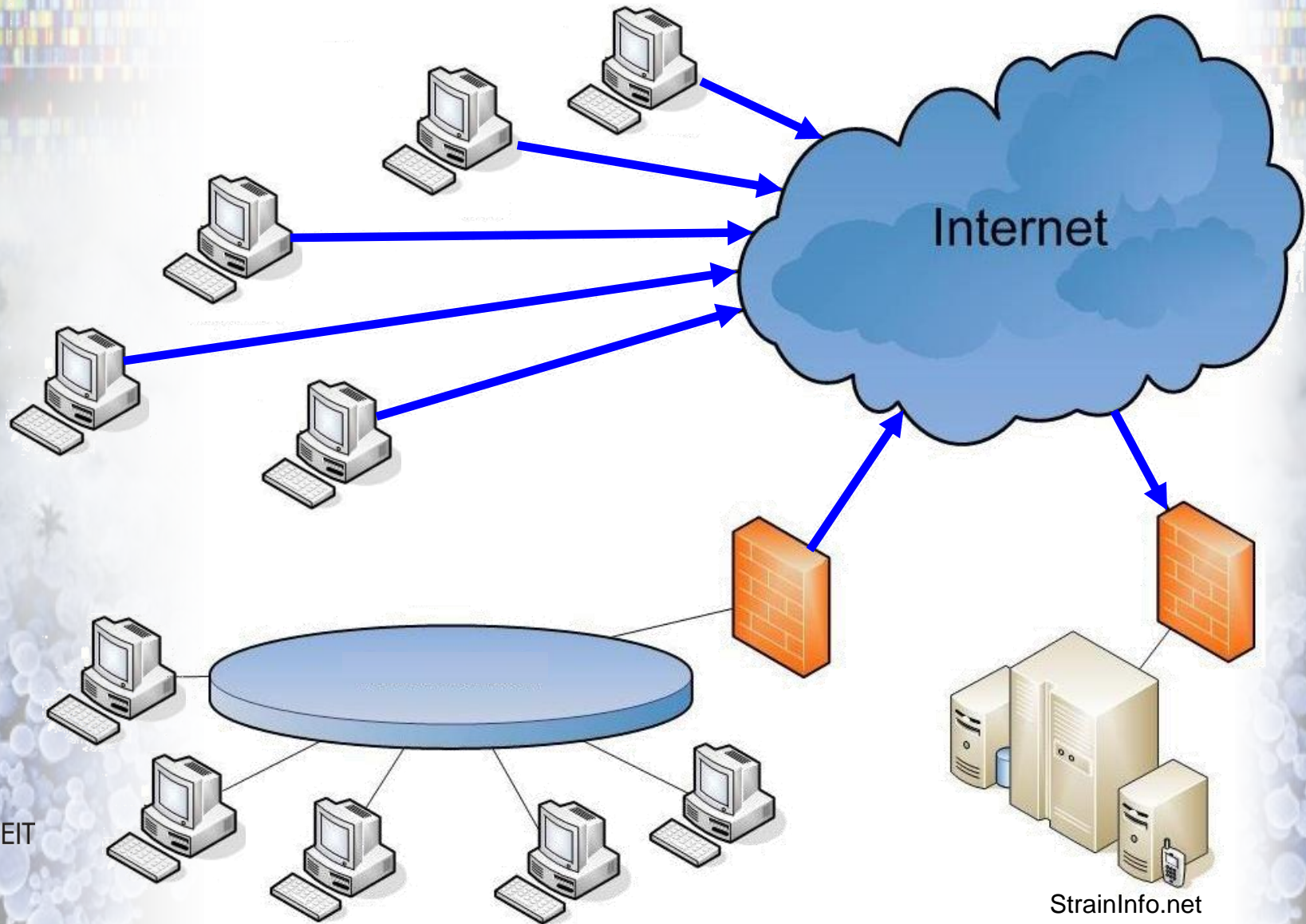
  <si:Medium>
    <si:mediumNumber>LMG Medium 185</si:mediumNumber>
    <si:mediumName>Bacteria Culture Medium 185</si:mediumName>
    <si:mediumURL>http://bccm.belspo.be/db/media_search_results.php?COLL=LMG&amp;FIELD=NUM&amp;TEXT1=185</si:mediumURL>
  </si:Medium>
  <si:growthTemperature>28</si:growthTemperature>
</si:Culture>
```

PUSH



StrainInfo.net
sync-server

Synchronization



Accumulative learning



Accumulative learning



publications

publish data

organisms

adopt to standards

community annotation

taxonomy

genomes

the sky is the limit

StrainInfo.net team

**Peter Dawyndt
Bernard De Baets
Paul De Vos
Wim De Smet
Wim Gillis
Bert Verslyppe**

**ALONG
CAME A
SPIDER**

WEB



www.StrainInfo.net