

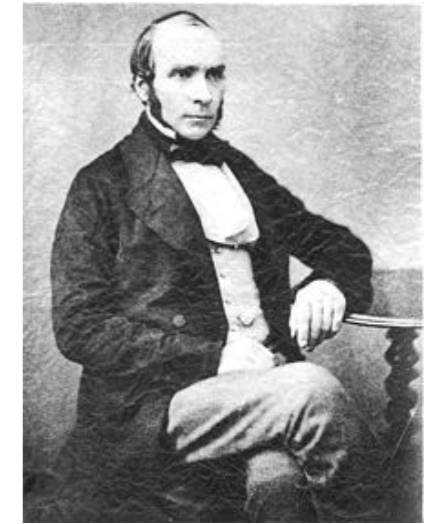
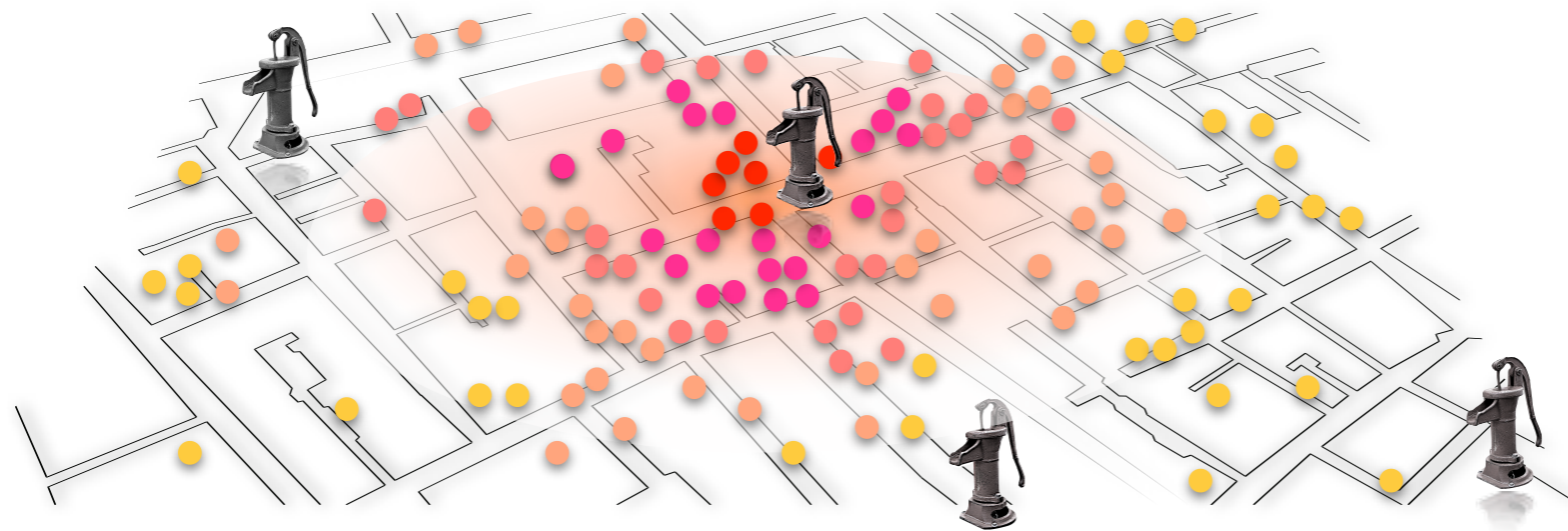


Open access *genetic data* for real-time epidemiology

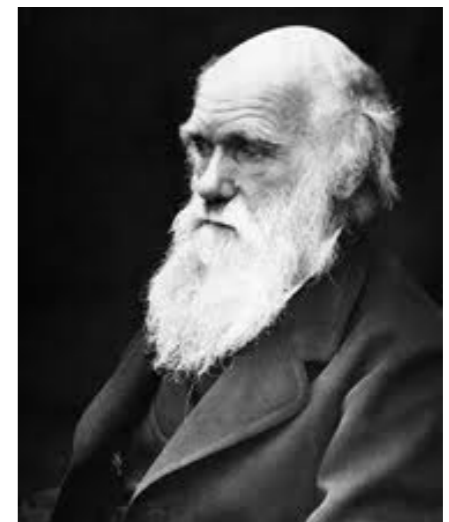
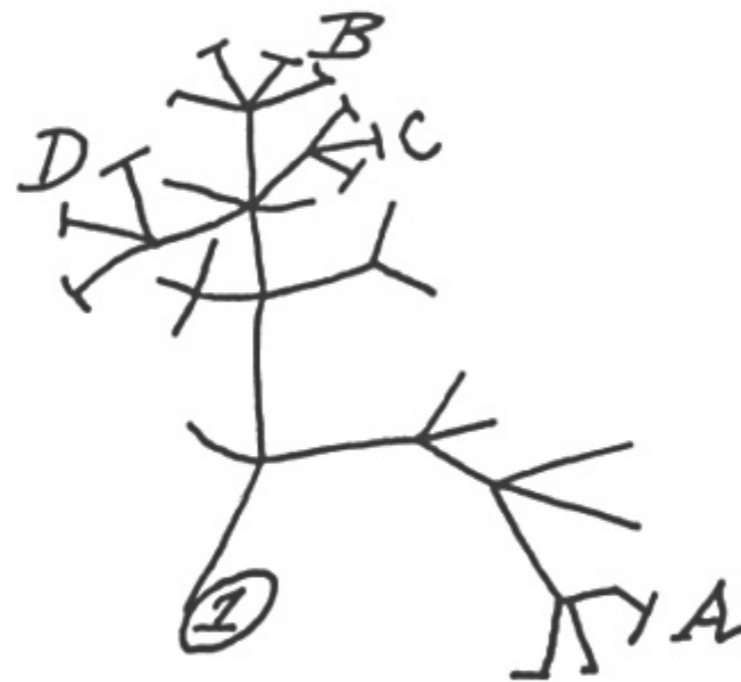
Philippe Lemey
Clinical and Evolutionary Virology
Department of Microbiology and Immunology
Rega Institute, KU Leuven, Belgium.
philippe.lemey@rega.kuleuven.be
rega.kuleuven.be/cev/ecv/



Molecular epidemiology of infectious disease

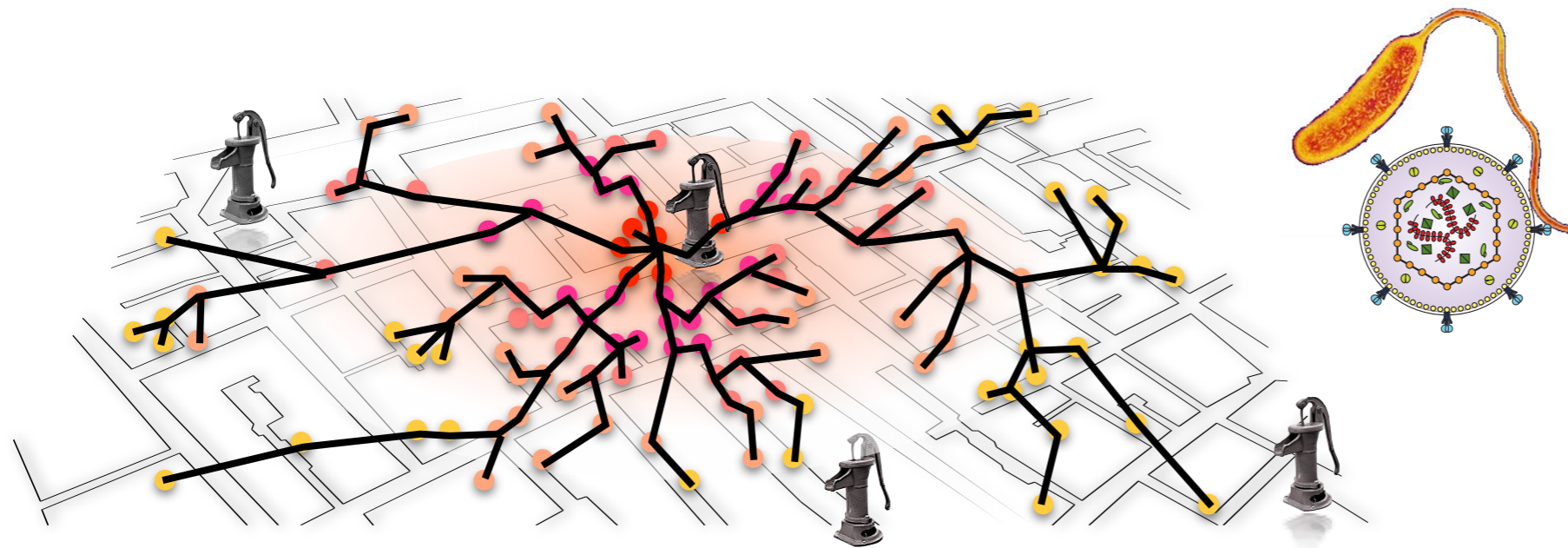


Snow J (1813-1858)

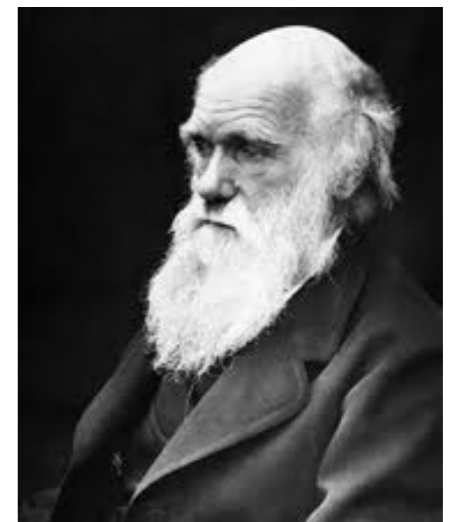
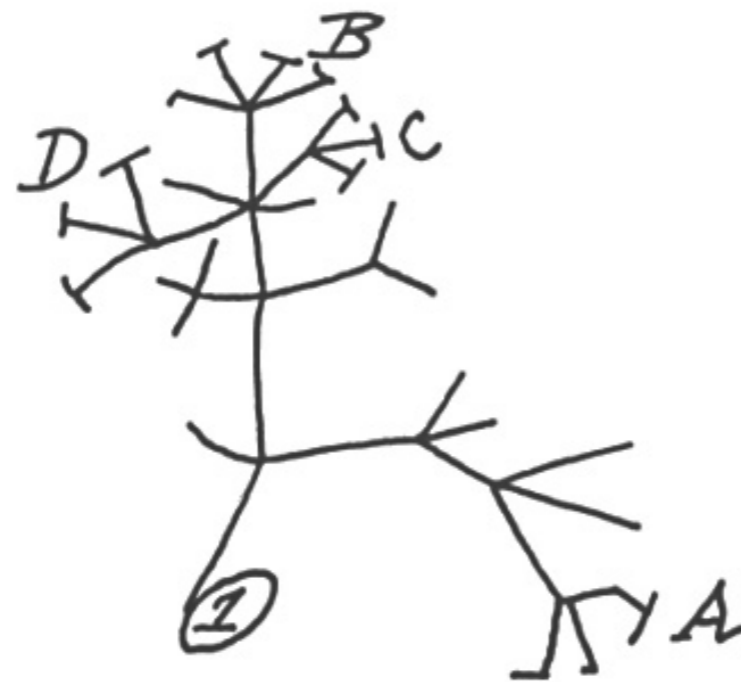


Darwin C (1809-1882)

Molecular epidemiology of infectious disease

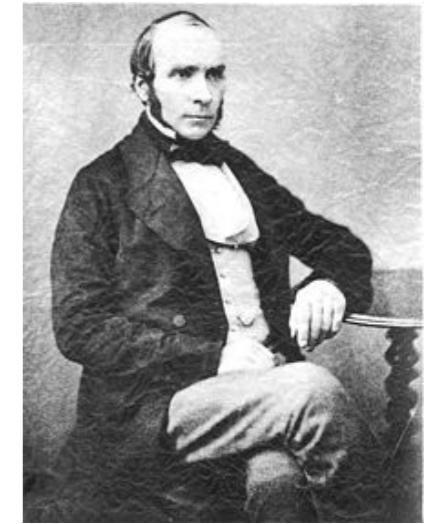
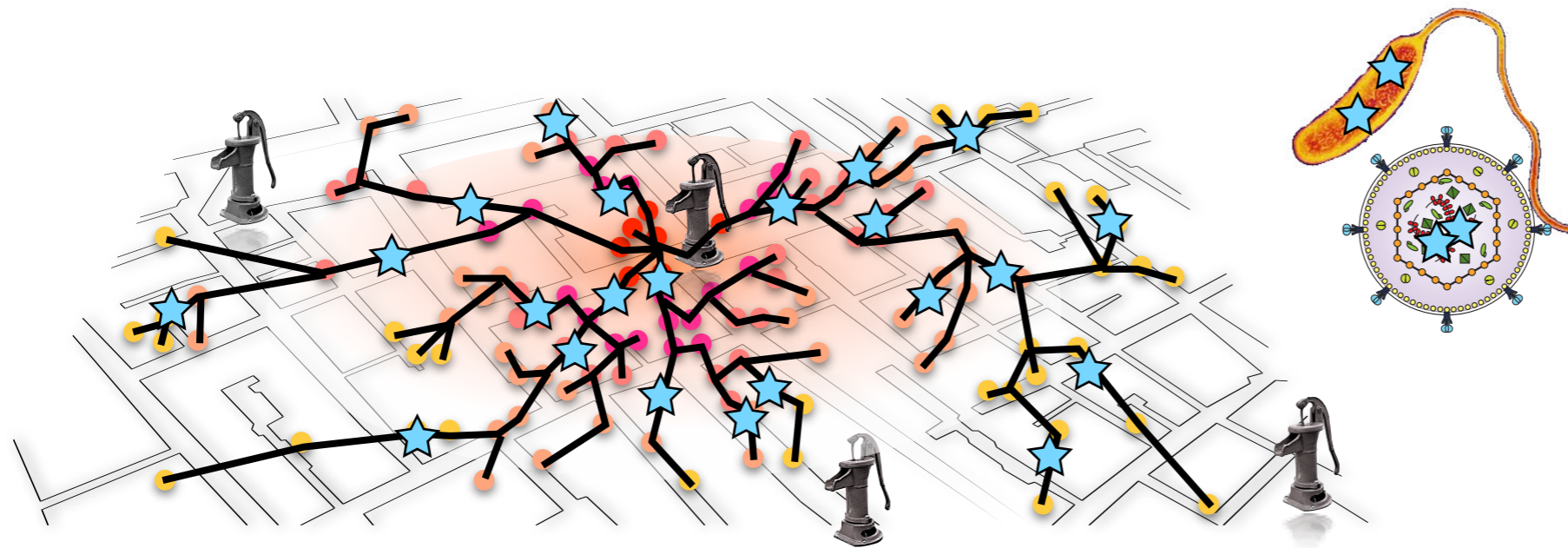


Snow J (1813-1858)

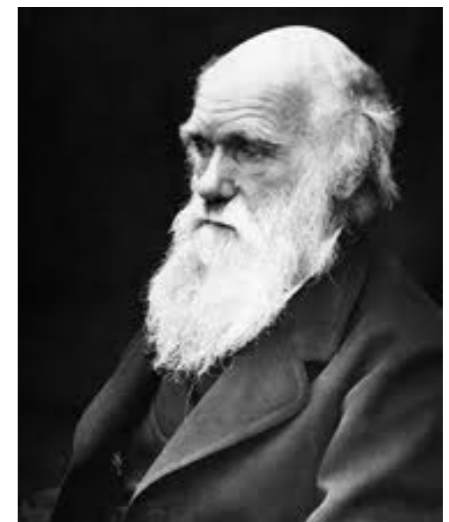
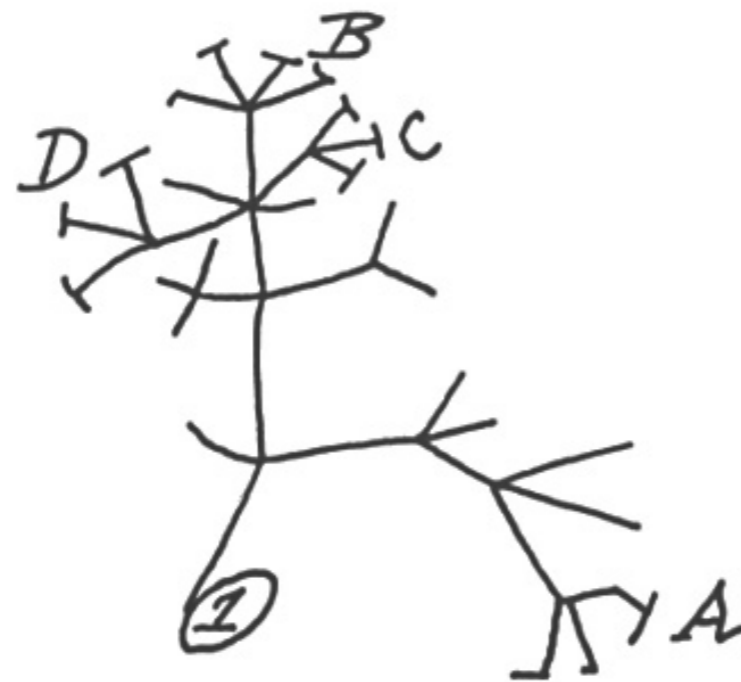


Darwin C (1809-1882)

Molecular epidemiology of infectious disease

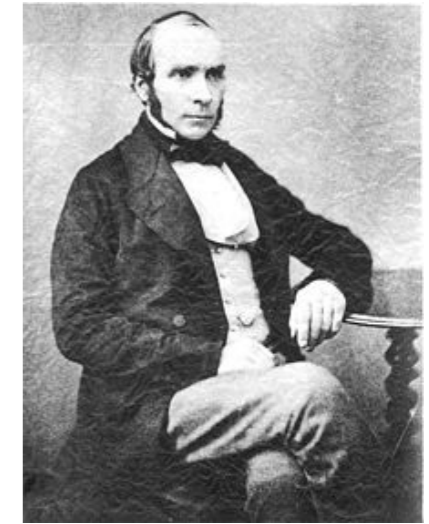
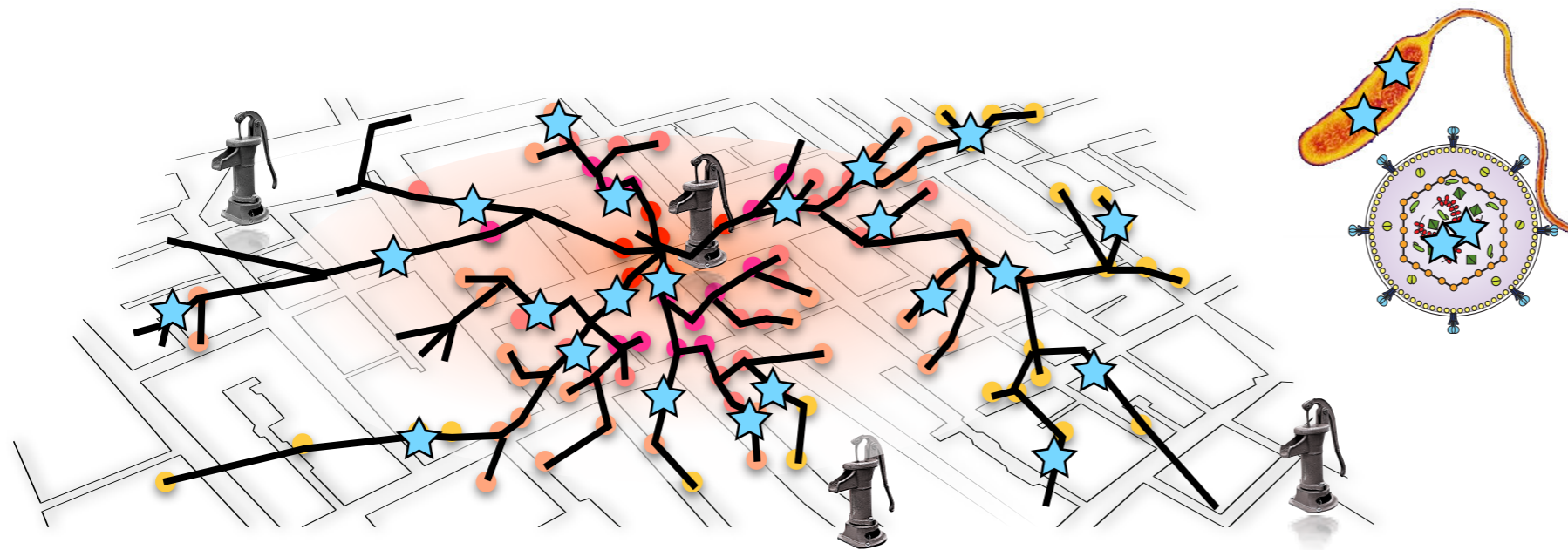


Snow J (1813-1858)

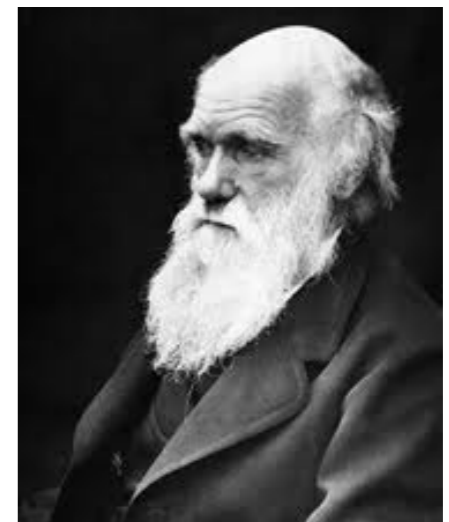
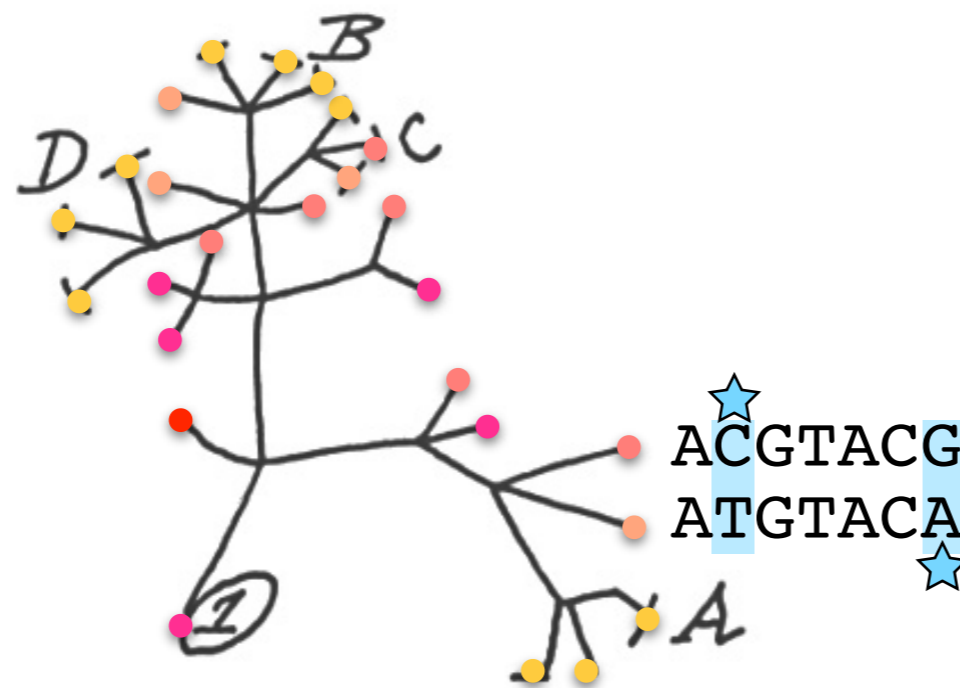


Darwin C (1809-1882)

Molecular epidemiology of infectious disease

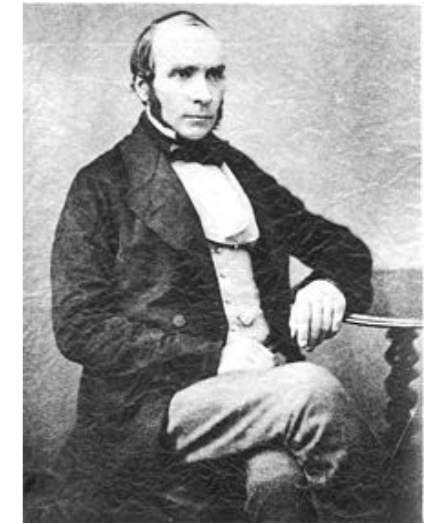
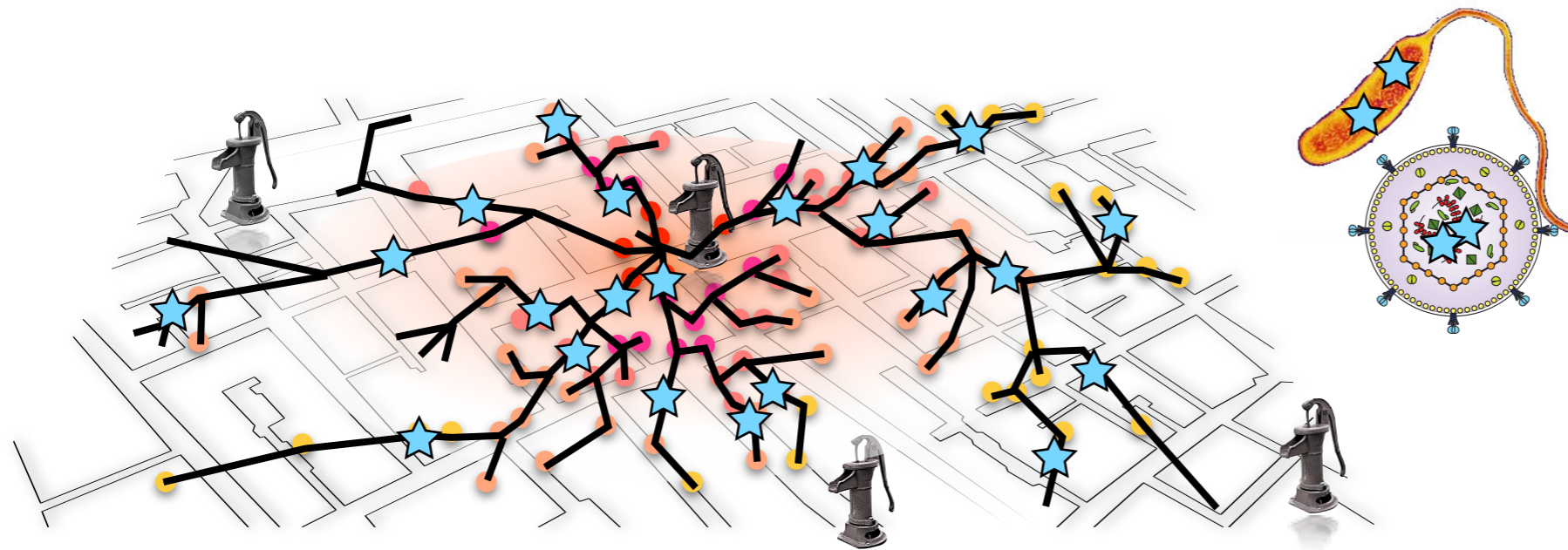


Snow J (1813-1858)

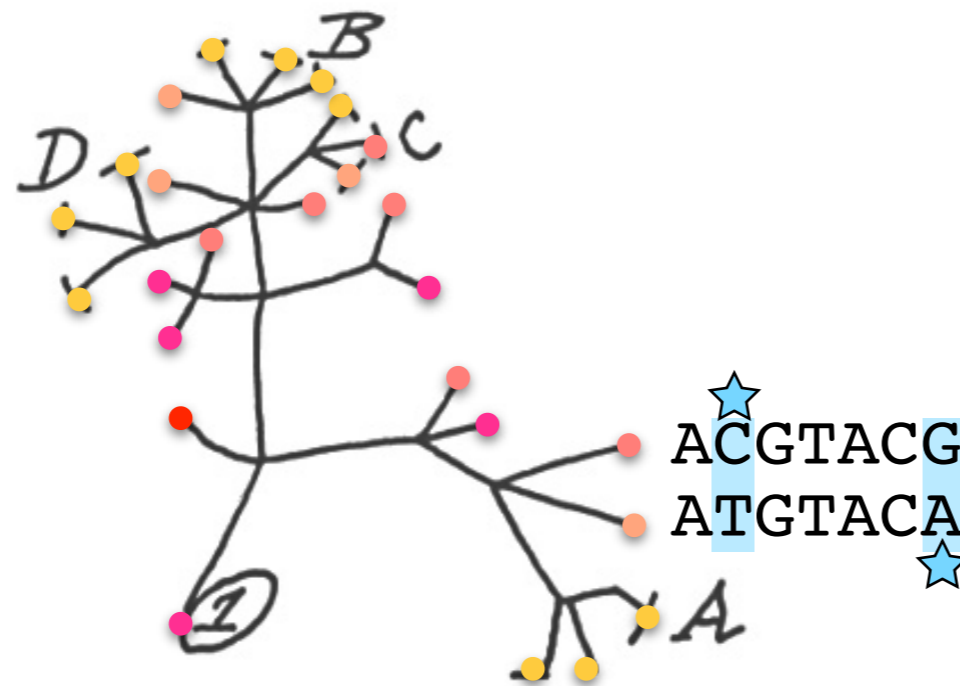


Darwin C (1809-1882)

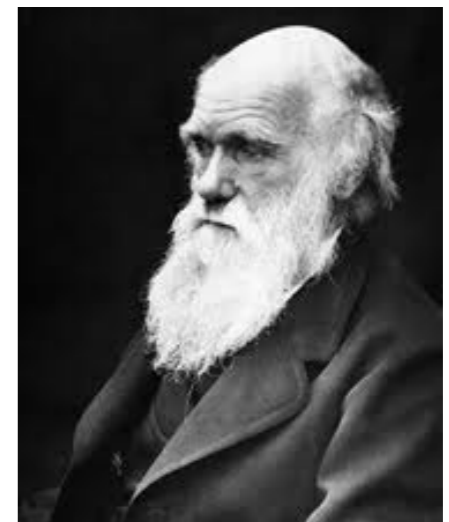
Molecular epidemiology of infectious disease



Snow J (1813-1858)

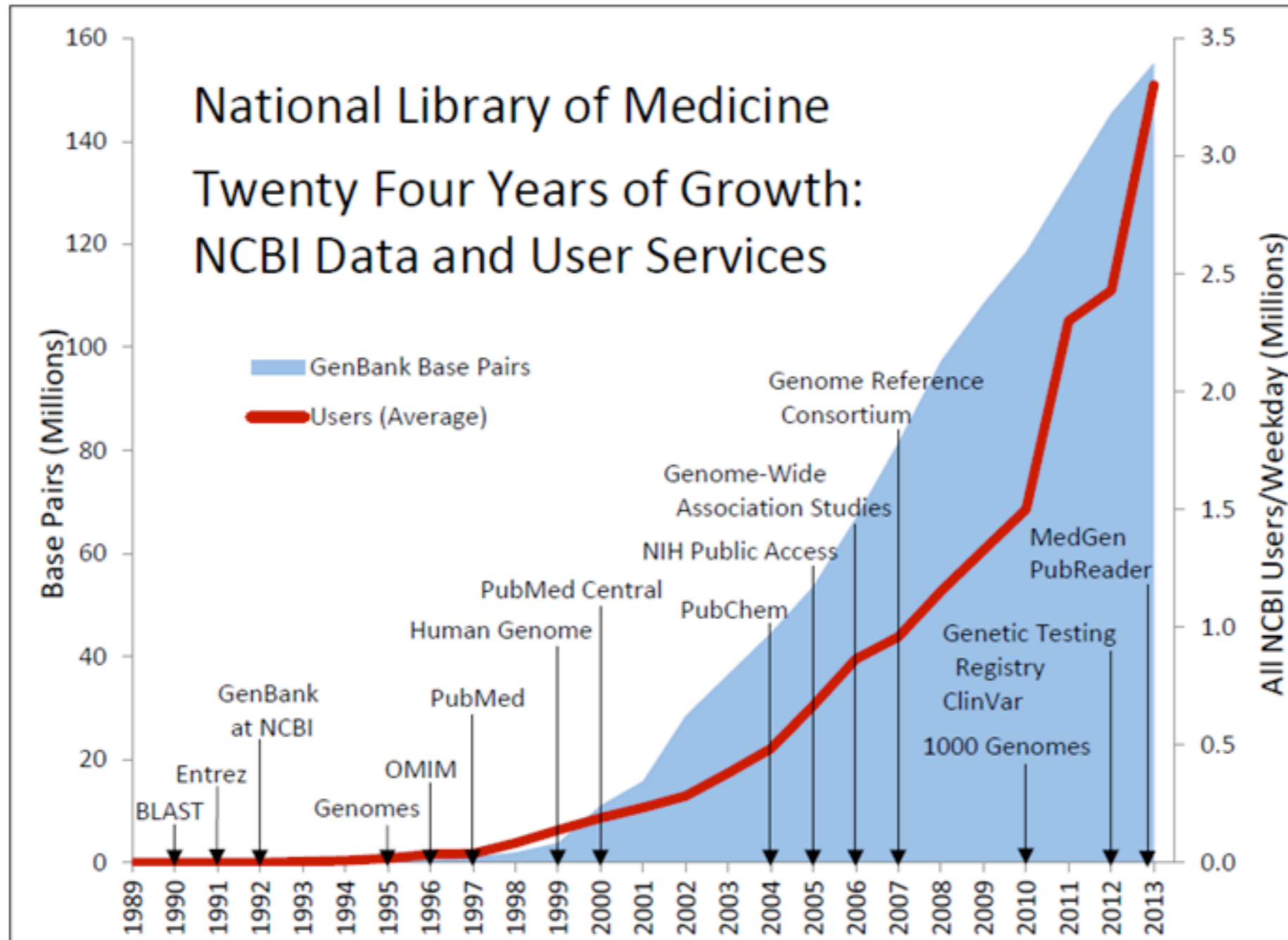


Phylodynamics and evolution



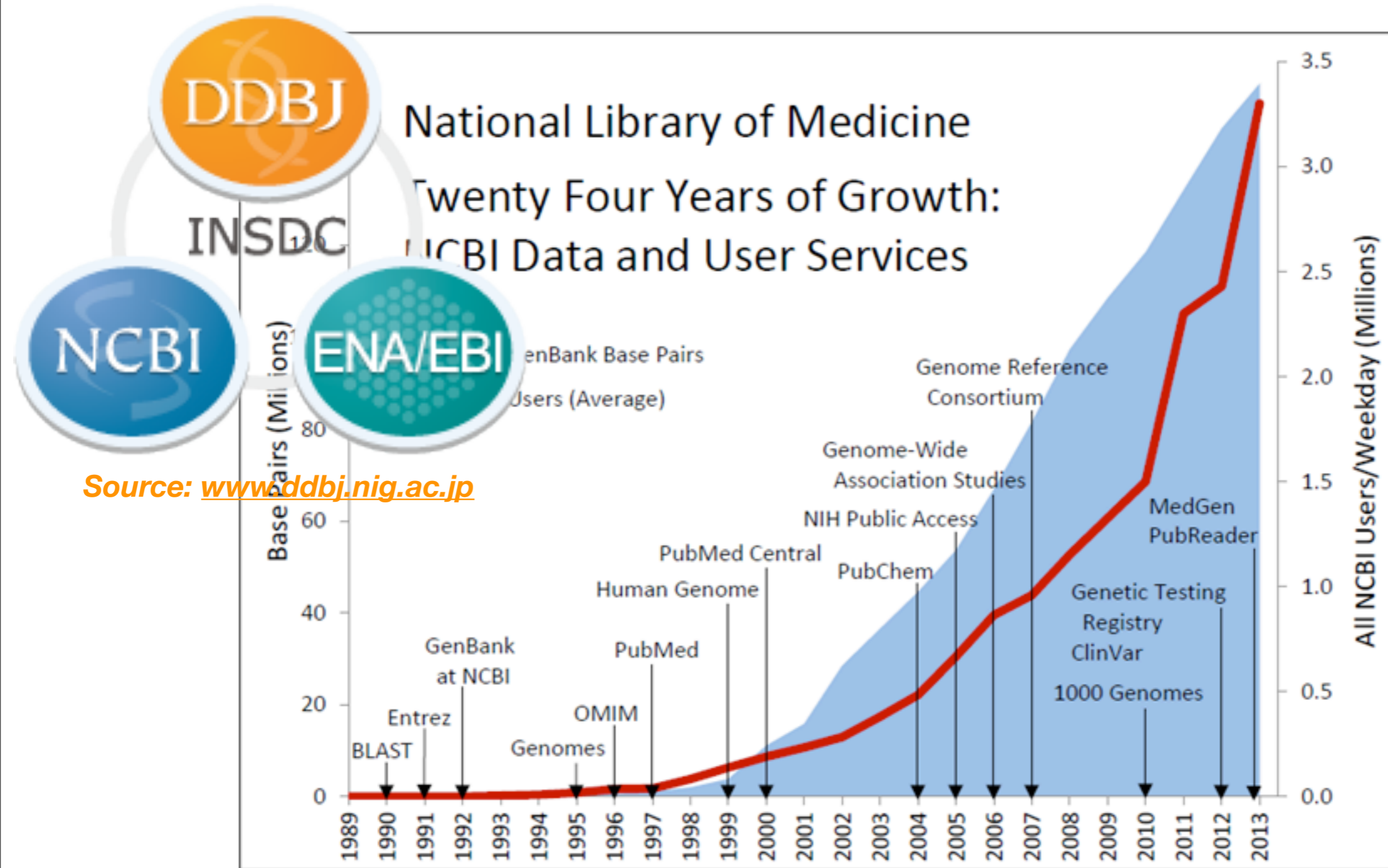
Darwin C (1809-1882)

The growth of public databases (GENBANK)



Source: the NNM 2015 Congressional Justification,
<http://www.nlm.nih.gov/about/2015CJ.html>

The growth of public databases (GENBANK)



Source: the NNM 2015 Congressional Justification,
<http://www.nlm.nih.gov/about/2015CJ.html>

Nucleic Acids Research

[ABOUT THIS JOURNAL](#) [CONTACT THIS JOURNAL](#) [SUBSCRIPTIONS](#)[CURRENT ISSUE](#) [ARCHIVE](#) [SEARCH](#)

2014 NAR Database Summary Paper

Source: http://www.oxfordjournals.org/our_journals/nar/database/cat/5

[Nucleotide Sequence Databases](#)[RNA sequence databases](#)[Protein sequence databases](#)[Structure Databases](#)[Genomics Databases \(non-vertebrate\)](#)[MGD - Mouse Genome Database](#)[The Gene Indices](#)[Genome annotation terms, ontologies and nomenclature](#)[Taxonomy and identification](#)[General genomics databases](#)[Viral genome databases](#)[CoVDB](#)[DPVweb](#)[euHCVdb](#)[HBVdb](#)[HCV Database](#)[HCVDB - Hepatitis C Virus Database](#)[HepSeq](#)[HERVd - Human Endogenous Retrovirus database](#)[HFV database](#)[HIV Drug Resistance Database](#)[HIV Interactions database](#)[HIV Molecular Immunology Database](#)[HIV RT and Protease Sequence Database](#)[HIV Sequence Database](#)[ISED](#)[IVDB - Influenza Virus Database](#)[NCBI Viral genomes](#)[Papillomavirus Episteme](#)[PhEVER](#)[phiSITE](#)[Poxvirus.org](#)[RNA Virus Database](#)[SARS-CoV RNA SSS](#)[Subviral RNA Database](#)[T4-like genome database](#)[The HIV Positive Selection Mutation Database](#)[VIDA](#)[VIPERdb](#)[ViPR](#)[Viral Bioinformatics Resource](#)[ViralORFeome](#)[ViralZone](#)[VirGen](#)[Virus Variation](#)[ViTa](#)

Nucleic Acids Research

ABOUT THIS JOURNAL CONTACT THIS JOURNAL SUBSCRIPTIONS CURRENT ISSUE ARCHIVE SEARCH

2014 NAR Database Summary Paper

Source: http://www.oxfordjournals.org/our_journals/nar/database/cat/5

- Nucleotide Sequence Databases
- RNA sequence databases
- Protein sequence databases
- Structure Databases
- Genomics Databases (non-vertebrate)
 - MGD - Mouse Genome Database
 - The Gene Indices
- Genome annotation terms, ontologies and nomenclature
- Taxonomy and identification
- General genomics databases
- Viral genome databases

- CoVDB
- DPVweb
- euHCVdb
- HBVdb
- HCV Database**
- HCVDB - Hepatitis C Virus Database
- HepSeq
- HERVd - Human Endogenous Retrovir
- HFV database
- HIV Drug Resistance Database
- HIV Interactions database
- HIV Molecular Immunology Database
- HIV RT and Protease Sequence Data

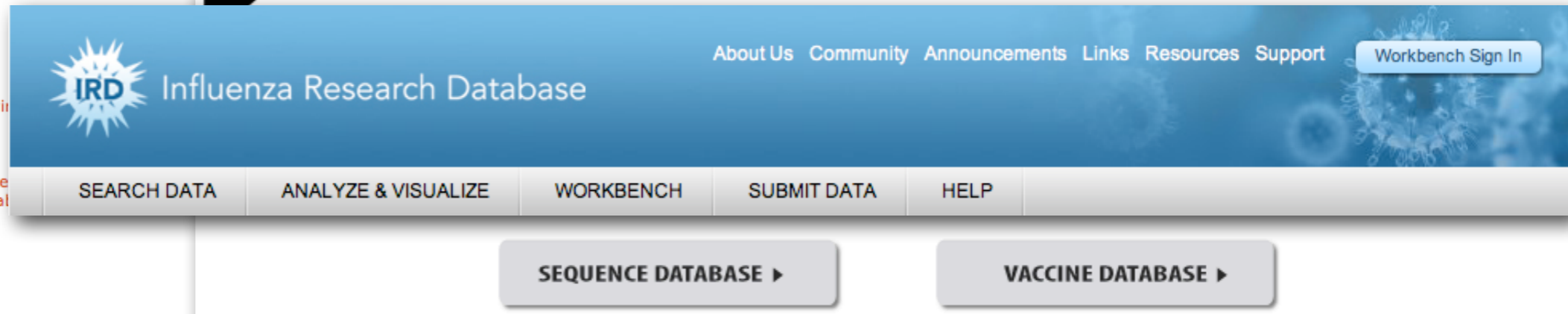
HIV Sequence Database

IVDB - Influenza Virus Database

- NCBI viral genomes
- Papillomavirus Episteme
- PhEVER
- phiSITE
- Poxvirus.org
- RNA Virus Database
- SARS-CoV RNA SSS
- Subviral RNA Database
- T4-like genome database
- The HIV Positive Selection Mutation
- VIDA
- VIPERdb
- ViPR**
- Viral Bioinformatics Resource
- ViralORFeome
- ViralZone
- VirGen
- Virus Variation
- ViTa



HIV DATABASES



IRD Influenza Research Database

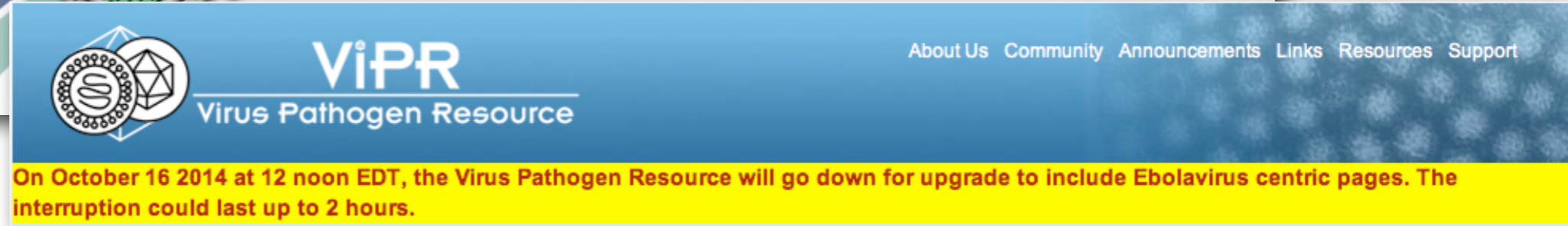
About Us Community Announcements Links Resources Support [Workbench Sign In](#)

SEARCH DATA ANALYZE & VISUALIZE WORKBENCH SUBMIT DATA HELP

SEQUENCE DATABASE ► VACCINE DATABASE ►



HCV DATABASES

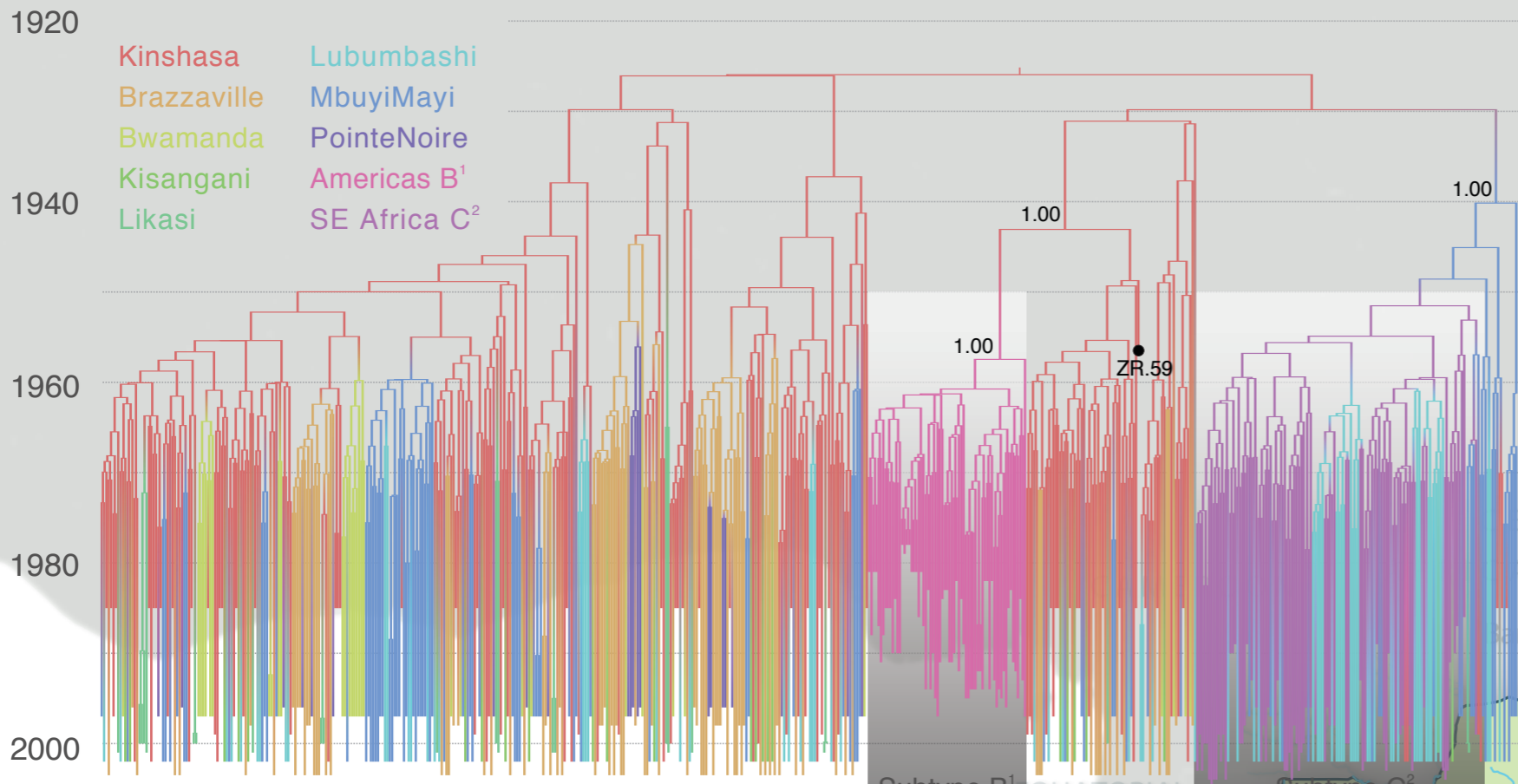


ViPR Virus Pathogen Resource

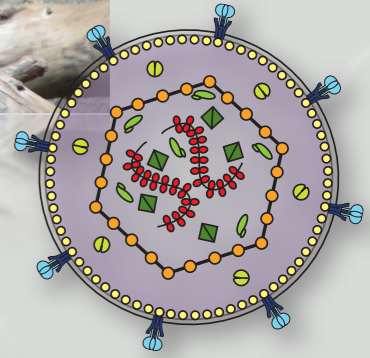
About Us Community Announcements Links Resources Support

On October 16 2014 at 12 noon EDT, the Virus Pathogen Resource will go down for upgrade to include Ebolavirus centric pages. The interruption could last up to 2 hours.

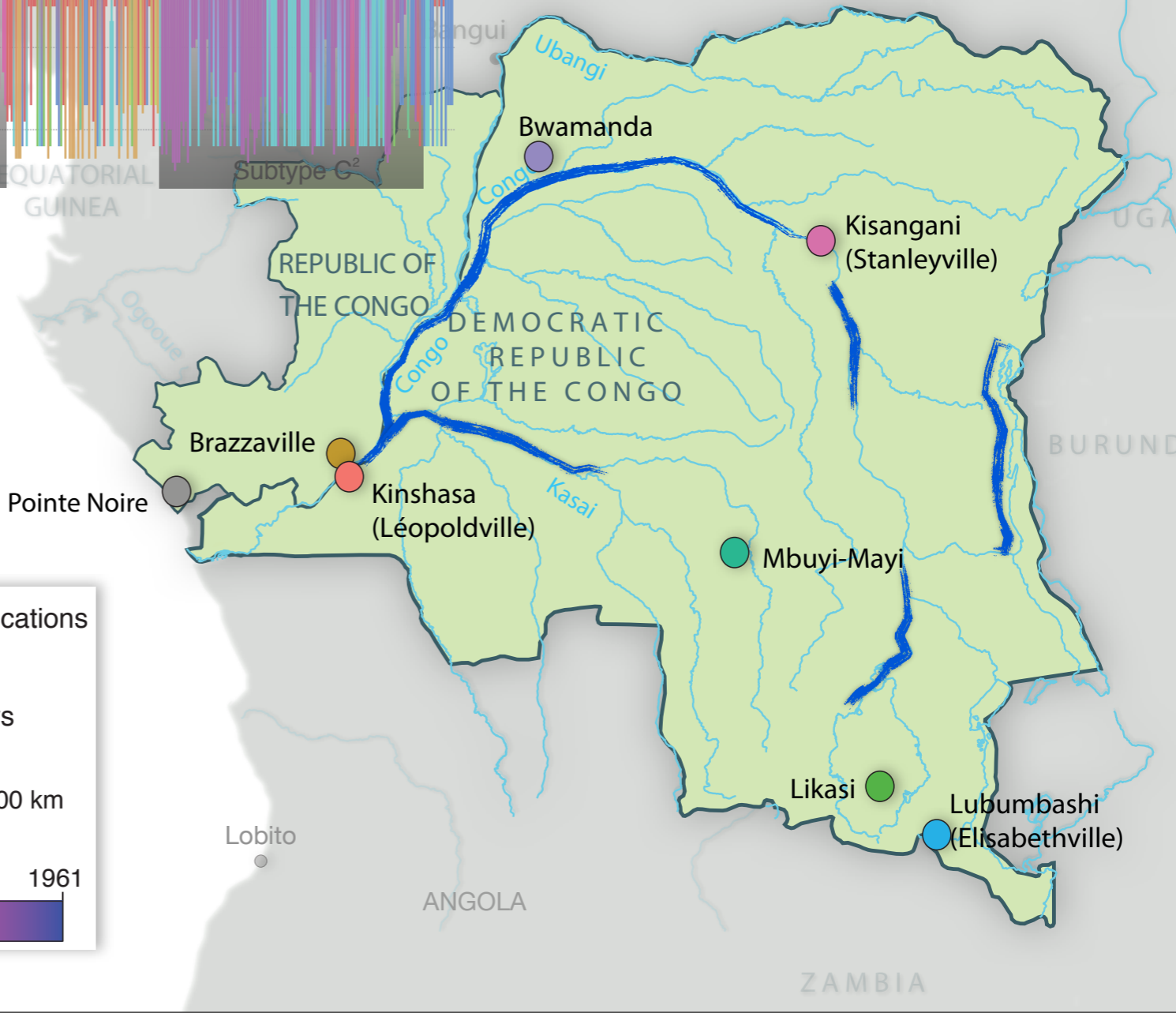




Source; Thomas Lersch



Faria et al., Science, 2014

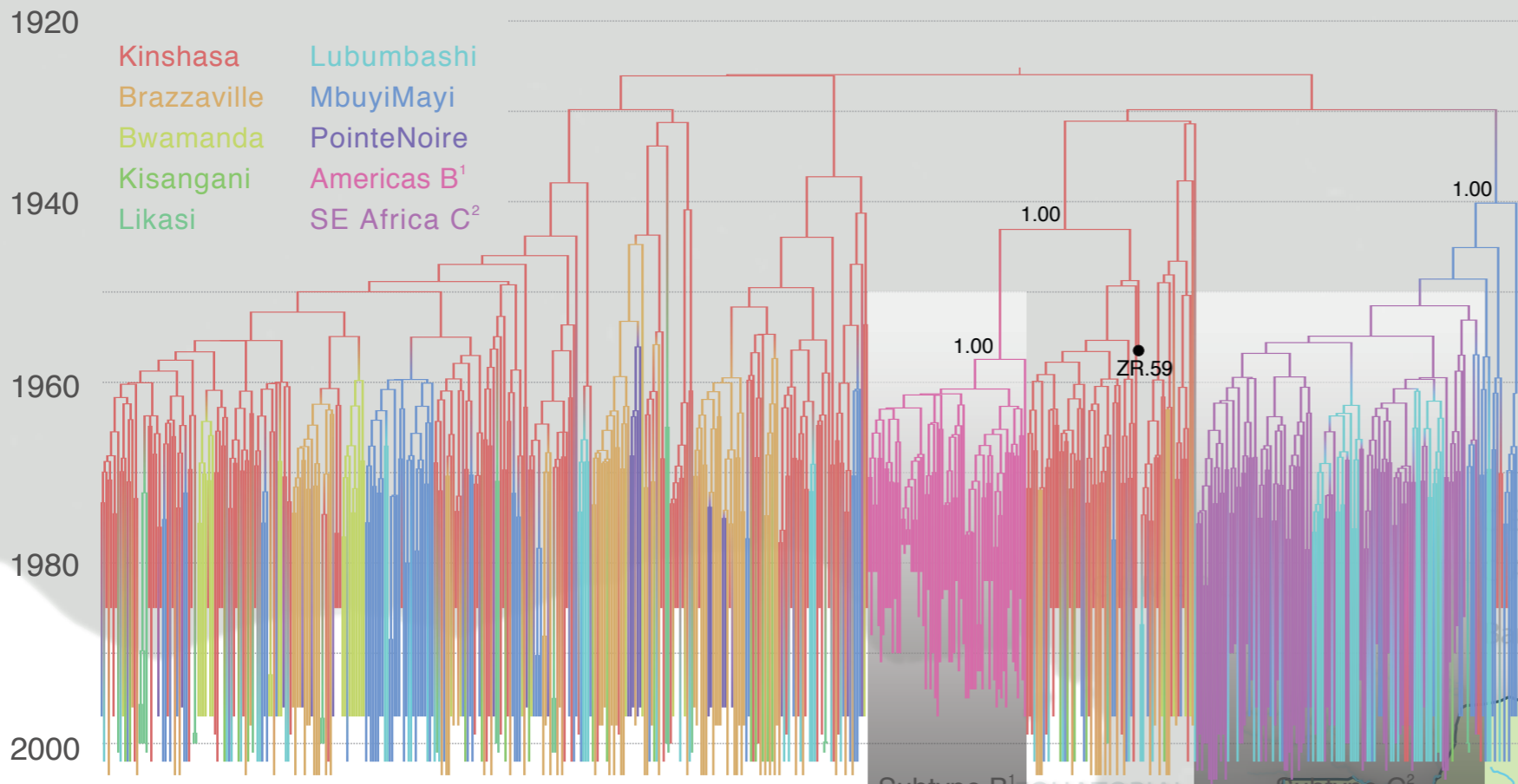


- Sample locations
- Railways
- Waterways

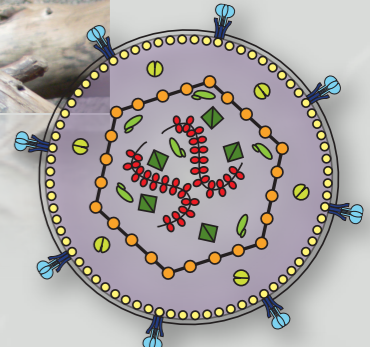
Spatial scale:
 200 km

Timescale:
 1920 1961

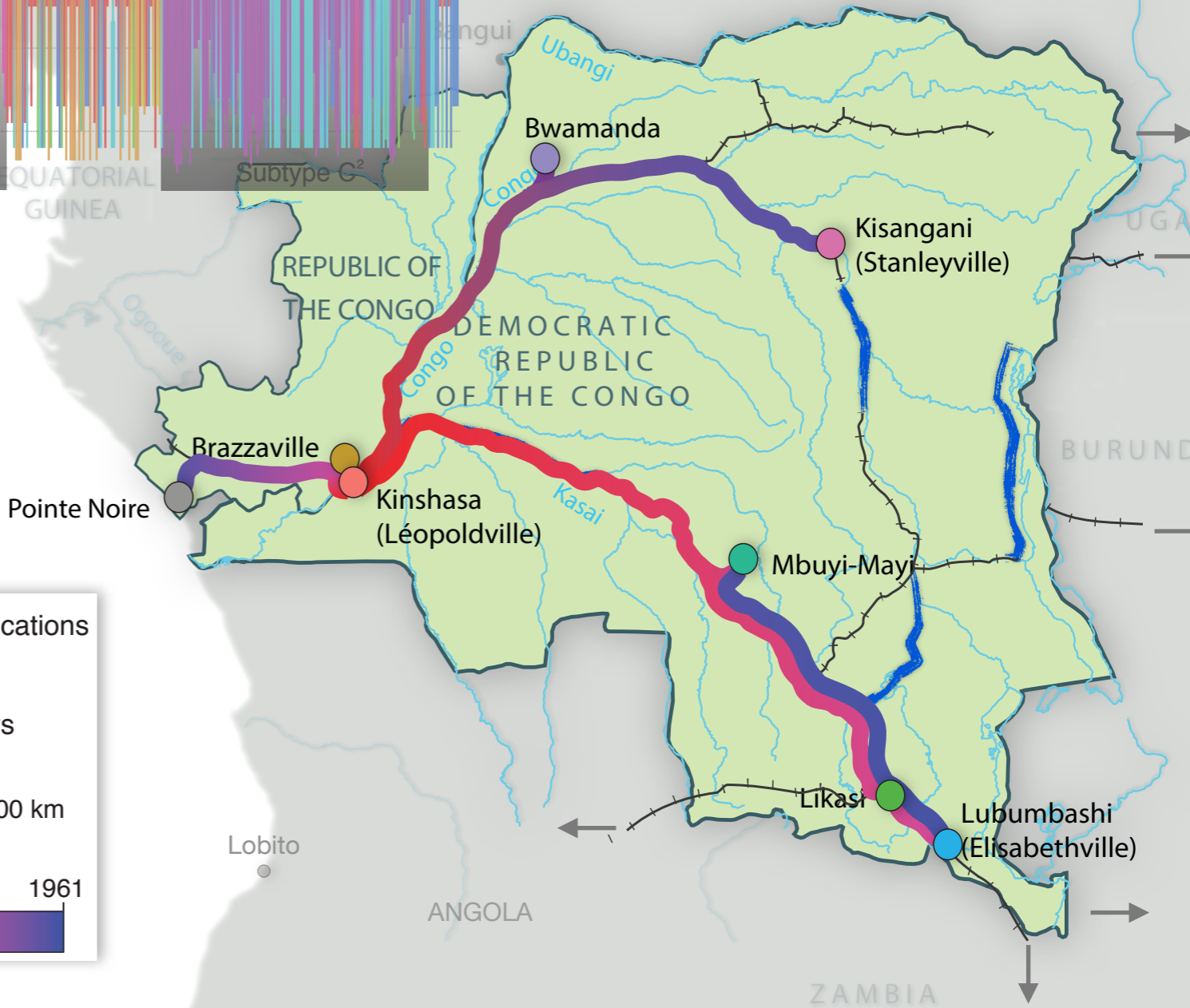




Source; Thomas Lersch



Faria et al., Science, 2014



- Sample locations
- Railways
- Waterways

Spatial scale:
 ─────────── 200 km

Timescale:
 1920 ─────────── 1961



Apocalypse Now

REEDUX



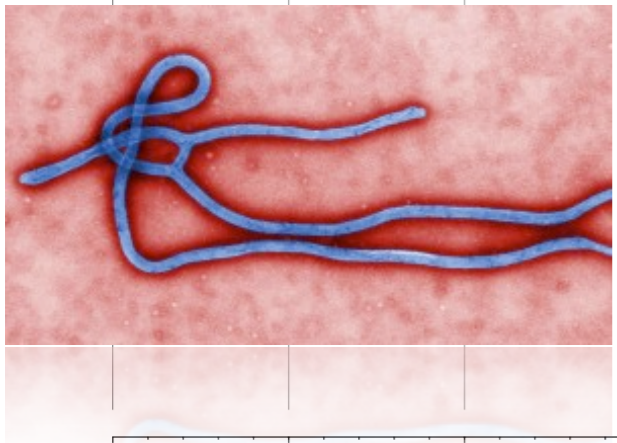
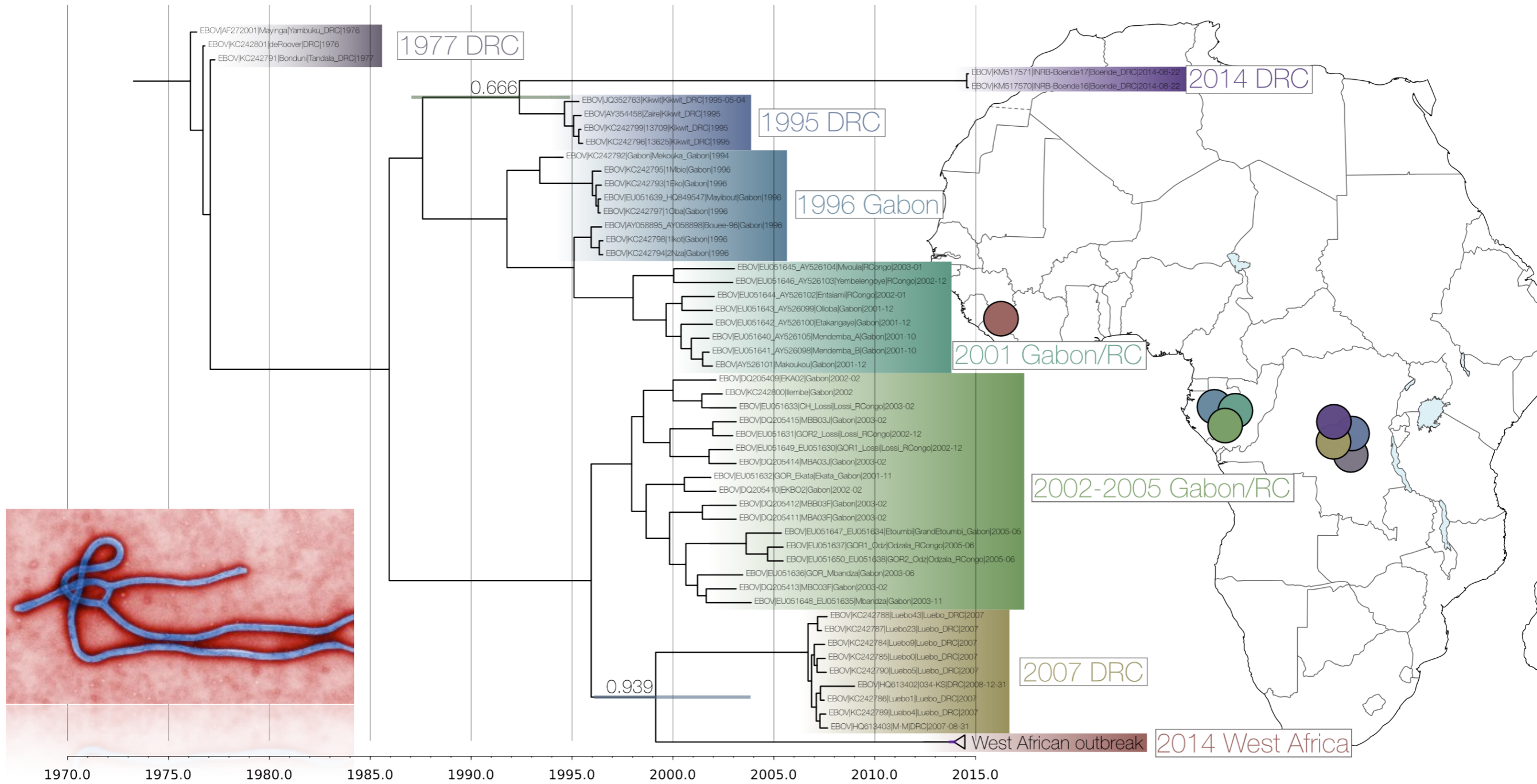
Photo Credits:<http://www.inquisitr.com/wp-content/2012/01/swine-flu-h1n1.jpg>

Apokalypse Now

REEDUX



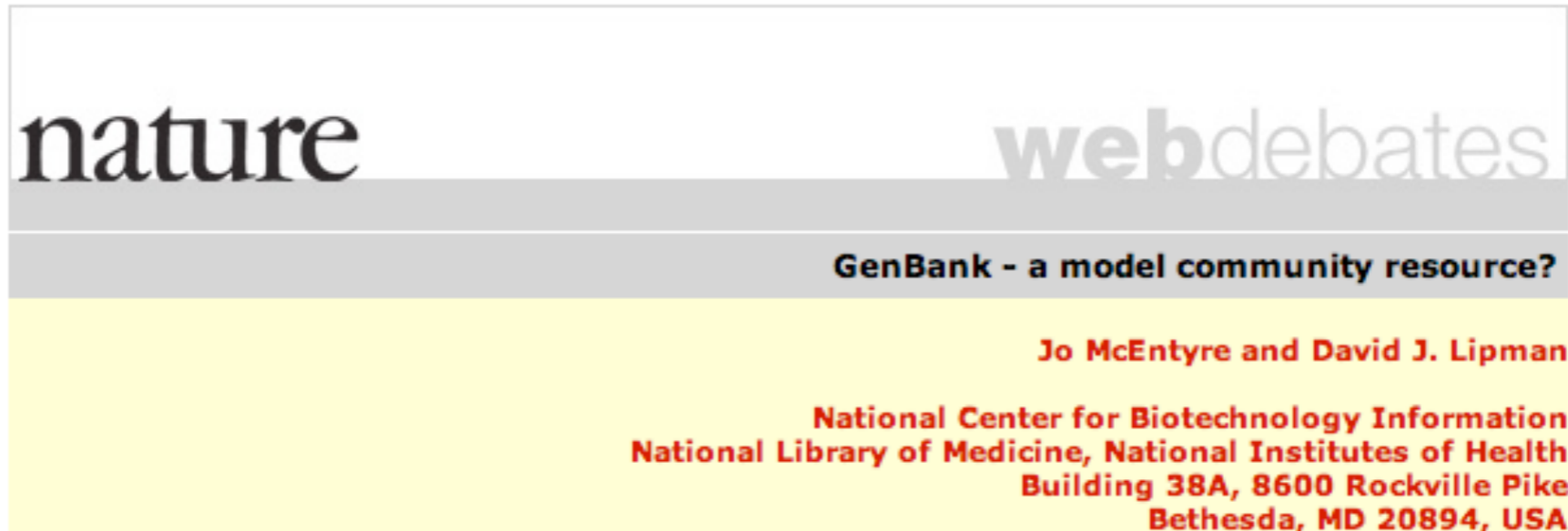
Photo Credits: <http://www.inquisitr.com/wp-content/2012/01/swine-flu-h1n1.jpg>



Gire et al., Science, 2014



A perfect world...



The image shows a screenshot of a web page header. On the left, the word "nature" is written in a large, black, serif font. To its right, the word "webdebates" is written in a smaller, grey, sans-serif font. Below these, a grey horizontal bar contains the text "GenBank - a model community resource?" in a bold, black, sans-serif font. Underneath this bar, a yellow background contains the authors' names "Jo McEntyre and David J. Lipman" in a bold, red, sans-serif font. Below the names, the affiliation "National Center for Biotechnology Information, National Library of Medicine, National Institutes of Health, Building 38A, 8600 Rockville Pike, Bethesda, MD 20894, USA" is written in a smaller, red, sans-serif font.

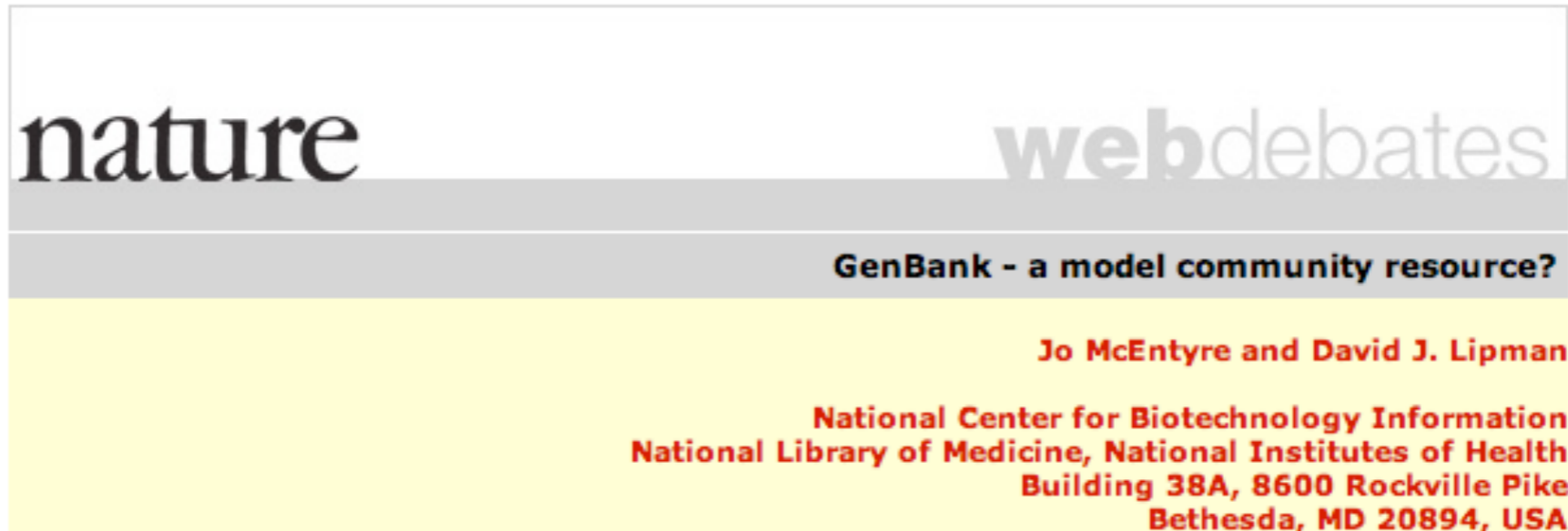
Genbank has become a household name among biologists. They all benefit from having free access to the 16 billion base pairs of primary DNA sequence and the related molecular information that has been submitted to this shared resource by the international scientific community. The information either goes directly to GenBank or is submitted via its counterparts in Europe -- the European Bioinformatics Institute in Cambridge (EBI) -- and Japan -- the DNA Data Bank of Japan (DDJB). GenBank demonstrates that, even in the fiercely competitive world of science, researchers recognize that contributing to large, shared data sets ultimately benefits everyone. The shared resource that is created is an indispensable tool that is greater than the sum of its parts.

Scientists have shown a willingness to place data in a community archive for the common good, knowing that it can be freely used by anyone. Moreover, all leading journals have adopted a policy that requires sequences to be deposited in the public databases, and the corresponding access numbers to be cited in published articles. All publicly funded laboratories now consider it *de rigueur* to contribute sequence data to Genbank within 24 hours of its generation, even if there is no accompanying research paper.

Source: <http://www.nature.com/nature/debates/e-access/Articles/lipman.html>



A perfect world, but...



The image shows a screenshot of a web page header. On the left, the word "nature" is written in a large, black, serif font. To its right, the word "webdebates" is written in a smaller, grey, sans-serif font. Below these, a grey horizontal bar contains the text "GenBank - a model community resource?" in a black, sans-serif font. Underneath the grey bar, a yellow background contains the authors' names "Jo McEntyre and David J. Lipman" in a red, sans-serif font. Below the names, the affiliation "National Center for Biotechnology Information, National Library of Medicine, National Institutes of Health, Building 38A, 8600 Rockville Pike, Bethesda, MD 20894, USA" is written in a red, sans-serif font.

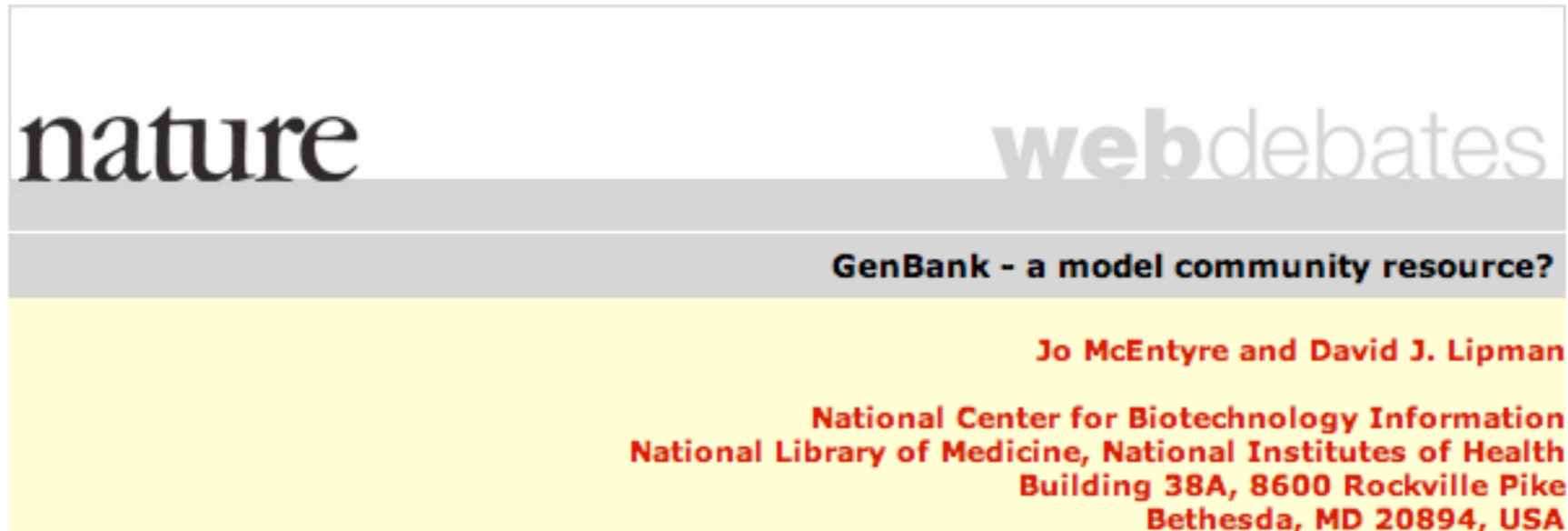
Genbank has become a household name among biologists. They all benefit from having free access to the 16 billion base pairs of primary DNA sequence and the related molecular information that has been submitted to this shared resource by the international scientific community. The information either goes directly to GenBank or is submitted via its counterparts in Europe -- the European Bioinformatics Institute in Cambridge (EBI) -- and Japan -- the DNA Data Bank of Japan (DDJB). GenBank demonstrates that, even in the ***fiercely competitive world of science***, researchers recognize that contributing to large, shared data sets ultimately benefits everyone. The shared resource that is created is an indispensable tool that is greater than the sum of its parts.

Scientists have shown a willingness to place data in a community archive for the common good, knowing that it can be freely used by anyone. Moreover, all leading journals have adopted a policy that requires sequences to be deposited in the public databases, and the corresponding access numbers to be cited in published articles. ***All publicly funded laboratories now consider it de rigueur to contribute sequence data to Genbank within 24 hours of its generation, even if there is no accompanying research paper.***

Source: <http://www.nature.com/nature/debates/e-access/Articles/lipman.html>



A perfect world, but...



The image shows a screenshot of a web page header. On the left, the word "nature" is written in a large, black, serif font. To its right, the word "webdebates" is written in a smaller, grey, sans-serif font. Below these, a grey horizontal bar contains the text "GenBank - a model community resource?" in a bold, black, sans-serif font. Underneath this bar, a yellow background contains the author's name "Jo McEntyre and David J. Lipman" in a bold, red, sans-serif font. Below the name, the affiliation "National Center for Biotechnology Information, National Library of Medicine, National Institutes of Health, Building 38A, 8600 Rockville Pike, Bethesda, MD 20894, USA" is written in a smaller, red, sans-serif font.

Genbank has become a household name among biologists. They all benefit from having free access to the 16 billion base pairs of primary DNA sequence and the related molecular information that has been submitted to this shared resource by the international scientific community. The information either goes directly to GenBank or is submitted via its counterparts in Europe -- the European Bioinformatics Institute in Cambridge (EBI) -- and Japan -- the DNA Data Bank of Japan (DDJB). GenBank demonstrates that, even in the ***fiercely competitive world of science***, researchers recognize that contributing to large, shared data sets ultimately benefits everyone. The shared resource that is created is an indispensable tool that is greater than the sum of its parts.

Scientists have shown a willingness to place data in a community archive for the common good, knowing that it can be freely used by anyone. ***Moreover, all leading journals have adopted a policy that requires sequences to be deposited in the public databases, and the corresponding access numbers to be cited in published articles.*** All publicly funded laboratories now consider it *de rigueur* to contribute sequence data to Genbank within 24 hours of its generation, even if there is no accompanying research paper.

Source: <http://www.nature.com/nature/debates/e-access/Articles/lipman.html>



Two steps forward one step back ... ?



platform.gisaid.org/

The GISAID Initiative is working around restrictions, which have previously prevented influenza information sharing, with the hope that more shared information will help researchers understand how viruses evolve, spread and potentially become pandemic. The access to GISAID is open to everyone, provided they identify themselves and agree to uphold the GISAID sharing mechanism.



Two steps forward one step back ... ?



platform.gisaid.org/

The GISAID Initiative is working around restrictions, which have previously prevented influenza information sharing, with the hope that more shared information will help researchers understand how viruses evolve, spread and potentially become pandemic. The access to GISAID is open to everyone, provided they identify themselves and agree to uphold the GISAID sharing mechanism.

Registration Form For Individual Users

The GISAID EpiFlu™ Database supports the underlying principles enshrined in GISAID, encourages the sharing of sequence and related data, and recognizes the contributions and interests of data providers and users by, for example:

- Requiring you to **acknowledge data contributors** , e.g. the Originating Laboratory where the clinical specimen or virus isolate was first obtained and the Submitting Laboratory where sequence data have been generated and submitted to the GISAID EpiFlu™ Database (a sample can be found [here](#)).



Acknowledging data contributors

We gratefully acknowledge the authors, originating and submitting laboratories of the sequences from GISAID's EpiFlu™ Database on which this research is based. The list is detailed below.

Segment ID	Segment	Country	Collection date	Isolate name	Originating Laboratory	Submitting Laboratory	Authors
EPI187670	HA	Japan	2009-Apr-02	A/AICHI/158/2009	Aichi Hospital	National Institute of Infectious Diseases, Tokyo	Odagiri.T
EPI187671	HA	Japan	2009-Apr-09	A/AICHI/161/2009	Aichi Hospital	National Institute of Infectious Diseases, Tokyo	Odagiri.T
EPI169724	HA	Japan	2008-Nov-26	A/AKITA/12/2008	Akita City Clinic	National Institute of Infectious Diseases, Tokyo	Odagiri.T
EPI187672	HA	Japan	2009-May-27	A/AKITA/34/2009	Akita City Clinic	National Institute of Infectious Diseases, Tokyo	Odagiri.T
EPI185826	HA	USA	2009-Feb-27	A/Arizona/08/2009	Arizona Medical Center	Centers for Disease Control & Prevention, Atlanta	Garten.R, Smith.C
EPI185829	HA	USA	2009-Feb-25	A/Arizona/11/2009	Arizona Medical Center	Centers for Disease Control & Prevention, Atlanta	Garten.R, Smith.C
EPI186310	HA	Australia	2009-Jan-12	A/Sydney/1/2009	Westmead Hospital	WHO Collaborating Centre for Reference and Research on Influenza, Melbourne	Deng.Y-M, Iannello.P, Caldwell.N, Leang.S-K, Komadina.N
EPI186311	HA	Australia	2009-Jan-20	A/Sydney/3/2009	Westmead Hospital	WHO Collaborating Centre for Reference and Research on Influenza, Melbourne	Deng.Y-M, Iannello.P, Caldwell.N, Leang.S-K, Komadina.N

All submitters of data may be contacted directly via the GISAID website www.gisaid.org

Two steps forward one step back ... ?



platform.gisaid.org/

The GISAID Initiative is working around restrictions, which have previously prevented influenza information sharing, with the hope that more shared information will help researchers understand how viruses evolve, spread and potentially become pandemic. The access to GISAID is open to everyone, provided they identify themselves and agree to uphold the GISAID sharing mechanism.

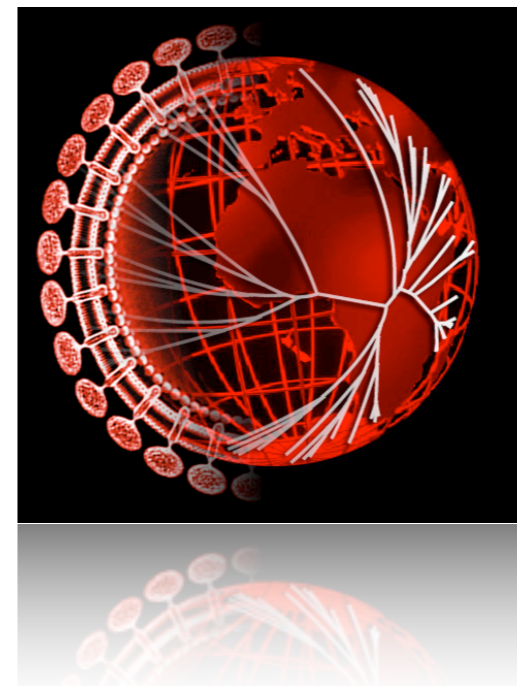
Registration Form For Individual Users

The GISAID EpiFlu™ Database supports the underlying principles enshrined in GISAID, encourages the sharing of sequence and related data, and recognizes the contributions and interests of data providers and users by, for example:

- Requiring you to **acknowledge data contributors** , e.g. the Originating Laboratory where the clinical specimen or virus isolate was first obtained and the Submitting Laboratory where sequence data have been generated and submitted to the GISAID EpiFlu™ Database (a sample can be found [here](#)).
- Preventing you from **attaching restrictions on the data** made available through the GISAID EpiFlu™ Database, such as including in a patent application any fraction of the sequence data obtained from GISAID. This will ensure that researchers will have unlimited access to the data for future generations to come;
- Requiring you to **make best efforts** to collaborate with the Originating Laboratory responsible for obtaining the specimens.
- Preventing you from **disclosing GISAID data outside the GISAID community** , such as by releasing GISAID sequences in any publication, transferring the data to colleagues that are not registered users, or offering GISAID data on a server accessible by others who are not duly registered with GISAID;



Recommendations and open questions



Recommendations and open questions

- *submit data to public database*
- *use unpublished genetic data in a collaborative spirit*
- *is the public domain sharing model not viable for genetic data?*
- *other credit systems for genetic data?*

