Cell

Genomics Goes Viral

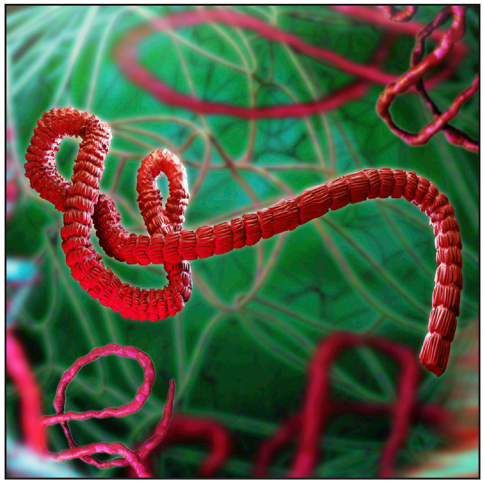
Epidemics are not a new event in human history. The Plague of Athens, one of the earliest documented outbreaks, hit the city state of Athens in ancient Greece during the Peloponnesian War and killed thousands of people. Bubonic Plague and Cholera decimated one-third of the European population over the middle and modern ages. The deadly H1N1 influenza pandemic of 1918 infected more than 500 million people and killed almost 5% of the world's population. As times change and science evolves, new tools to treat and contain infectious disease arise. On the one hand, antibiotics, antivirals, and vaccines are routinely available. On the other, urbanization, social dynamic changes, and high mobility allow for ignition and spread of new pathogens at a much higher rate. How does an epidemic begin? Where it comes from? How does it spread? Perhaps as relevant, how can it be stopped? Recent studies reveal the power of using genomic data to model and reconstruct the evolution of the two major threats to human health: the global HIV-1 epidemics and the very recent outbreak of Ebola virus in West Africa.

Interspecies transmission of HIV-1 might have first occurred in southeastern Cameroon. Fluvial connections between southern Cameroon and Kinshasa were frequent by that time, and the virus could have travelled through its host to Kinshasa, where it found fertile ground for further spread.

It has been 30 years since HIV-1 was found to be the causative agent of AIDS (UNAIDS 2013). Still, despite all of the information gathered about the genetics of the virus, as well as its simian counterpart, SIV, the initial steps that allowed the establishment of the virus in human populations remain enigmatic. It is also illusory why the HIV-1 group M (for "major") was the only to reach pandemic proportions, whereas other groups such as O (for "outlier"), N, and P remain contained in Central Africa. Now, a group led by Oliver Pybus and Philippe Lemey attempts to understand the early dynamic of the epidemics, using genomic sequence data already available in the HIV Sequence Database maintained by the Los Alamos National Laboratory in New Mexico (Faria et al., 2014). They analyzed hundreds of samples from group M viruses from different African locations over the past century, beginning in the region formerly known as Belgian Congo, then Zaire and now Democratic Republic of Congo (DRC), and extended their analysis to neighbor areas. The sequencing data was assembled to create phylogenetic trees, and a technique known as "molecular clock" was used to establish when each tree originated and when it produced new branches.

They date the ignition of the HIV epidemics as early as 1920 and support the suggestion that the first episode of transmission from chimpanzees to humans happened in the surrounds of Kinshasa in the early 1900s. The active transportation network available in Africa in the early 20th century, associated with changes in sexual behavior and iatrogenic medical interventions, played a major role in creating the perfect environment for further dissemination of the virus, which reached Lubumbashi (then Elizabethville) in 1937 and Mbuji-Mayi in 1939—major areas of mining that were well served with rail lines. The HIV-1 group M, subtype B, which accounts for the majority of infections in the United States and Europe, likely surfaced in Kinshasa by 1944. Interestingly, they find that HIV-1 groups M and O expanded at similar rates until 1960. Though it is still unclear why group M outpaced group O and became predominant, this study brings our understanding of the early events in the HIV epidemics to a much higher level of resolution and provides a perspective on how social and economic changes are big players in the emergence of new infectious agents.

Reconstructing the history of an epidemic that began almost a century ago has its own challenges. What about using genomic surveillance to understand an outbreak as it unveils? The 2014 outbreak of the deadly Ebola virus began in February in Guinea and quickly spread through Guinea, Liberia, Sierra Leone, and Nigeria (Baize et al., 2014). It is the largest Ebola outbreak ever known, and first cases in the United States and Europe, imported from Africa, are now confirmed. To understand the origins of the virus and its transmission chain in the current outbreak, a multinational team led by Pardis Sabeti at Harvard University sequenced the genome of the Ebola virus from samples of 78 patients from Sierra Leone who died or were infected with the 2014 virus, combined their data with previously published libraries, and reconstructed the spread of the outbreak (Gire et al., 2014).



Ebola virus, formerly called Zaire ebolavirus, is one of the four viruses of the genus Ebolavirus that can produce disease in humans, causing a type of hemorrhagic fever with very high fatality rate.

The analysis suggests that a single event of transmission from an Ebola natural reservoir into a human triggered the whole process. Molecular dating places the ancestor virus of the sequenced isolates around late February 2014, near when the first cases were confirmed in Guinea. Epidemiological investigations traced the initial cases in Sierra Leone to a burial ceremony of a traditional healer who had treated patients with Ebola in Guinea. Two genetically distinct viruses circulating in Guinea were introduced in Sierra Leone simultaneously in late May 2014. All subsequent viral diversity in Sierra Leone accumulated from the two original viral backbones, showing that, in opposition to previous Ebola outbreaks, continued human-reservoir exposure did not contribute to growth of the epidemics. As always happens with Ebola virus, the new outbreak is accompanied by extensive genetic changes in the virus. Whether these new mutations contribute to high lethality observed in this epidemics and whether additional mutations have emerged since the first Sierra Leonean isolates remains to be determined. Further studies will require access and sequencing of more samples from later stages of the epidemics.

Studies such as these two might not be able to prevent future outbreaks. However, the ability to gather genomic data and combine it with epidemiological vigilance and modeling might teach us how to prepare and perhaps how to interrupt the chain of events before it is too late.

REFERENCES AND RECENT RELATED PAPERS

Baize, S., Pannetier, D., Oestereich, L., Rieger, T., Koivogui, L., Magassouba, N., Soropogui, B., Sow, M.S., Keïta, S., De Clerck, H., et al. (2014). N. Engl. J. Med. *371*, 1418–1425.

Faria, N.R., Rambaut, A., Suchard, M.A., Baele, G., Bedford, T., Ward, M.J., Tatem, A.J., Sousa, J.D., Arinaminpathy, N., Pépin, J., et al. (2014). Science *346*, 56–61.

Gire, S.K., Goba, A., Andersen, K.G., Sealfon, R.S., Park, D.J., Kanneh, L., Jalloh, S., Momoh, M., Fullah, M., Dudas, G., et al. (2014). Science *345*, 1369–1372.

UNAIDS (2013) http://www.unaids.org/en/resources/campaigns/globalreport2013/globalreport/.

João Monteiro