

# Unexpected cross-species contamination in genome sequencing projects

November 18 2014

---

As genome sequencing has gotten faster and cheaper, the pace of whole-genome sequencing has accelerated, dramatically increasing the number of genomes deposited in public archives. Although these genomes are a valuable resource, problems can arise when researchers misapply computational methods to assemble them, or accidentally introduce unnoticed contaminations during sequencing.

The first complete bacterial genome, *Haemophilus influenzae*, appeared in 1995, and today the public GenBank database contains over 27,000 prokaryotic and 1,600 eukaryotic genomes. The vast majority of these are draft genomes that contain gaps in their sequences, and researchers often use these draft sequences for future analyses.

Each [genome sequencing](#) project begins with a DNA source, which varies depending on the species. For animals, blood is a common source, while for smaller organisms such as insects the entire organism or a population of organisms may be required to yield enough DNA for sequencing. Throughout the process of DNA isolation and sequencing, contamination remains a possibility. Computational filters applied to the raw sequencing reads are usually effective at removing common laboratory contaminants such as *E. coli*, but other contaminants may be more difficult to identify.

In a new study in *PeerJ*, authors from Johns Hopkins University discovered contaminating bacterial and viral sequences in "draft" assemblies of animal and [plant genomes](#) that had been deposited in

GenBank. These may cause particular problems for the rapidly growing field of microbiome analysis, when sequences labeled as animal in origin actually turn out to be microbial.

In an even more surprising finding, the authors discovered the presence of cow and sheep DNA in the supposedly finished genome of a pathogenic bacterium, *Neisseria gonorrhoeae*. Although deposited in GenBank as a finished genome, the bacterium apparently was a [draft genome](#) that was submitted as complete, with erroneous DNA inserted in five places. If taken at face value, this data would appear to be a startling case of lateral gene transfer, but the correct explanation appears to be more mundane.

These findings highlight the importance of careful screening of DNA sequence data both at the time of release and, in some cases, for many years after publication.

Provided by PeerJ

Citation: Unexpected cross-species contamination in genome sequencing projects (2014, November 18) retrieved 24 April 2024 from <https://phys.org/news/2014-11-unexpected-cross-species-contamination-genome-sequencing.html>

|  |
|--|
| <p>This document is subject to copyright. Apart from any fair dealing for the purpose of private study or research, no part may be reproduced without the written permission. The content is provided for information purposes only.</p> |
|--|