# REPORT DOCUMENTATION PAGE

Form Approved OMB NO. 0704-0188



| 1. REPORT DATE (DD-MM-YYYY) | 2. REPORT TYPE | 3. DATES COVERED (From - To) |
|---|---|---|
| 31-01-2014 | Final Report | 1-Sep-2010 - 31-Oct-2013 |

4. TITLE AND SUBTITLE
Continuation of Novel Analytical and Empircal Approaches to the Origin and Prediction of Pathogenicity

5a. CONTRACT NUMBER
W911NF-10-1-0339

5b. GRANT NUMBER

5c. PROGRAM ELEMENT NUMBER
9D10U6

6. AUTHORS
Ward Wheeler, Daniel Janies, Joel Cracraft

5d. PROJECT NUMBER

5e. TASK NUMBER

5f. WORK UNIT NUMBER

7. PERFORMING ORGANIZATION NAMES AND ADDRESSES
American Museum of Natural History
Central Park West at 79th Street

New York, NY 10024 -5193

8. PERFORMING ORGANIZATION REPORT NUMBER

9. SPONSORING/MONITORING AGENCY NAME(S) AND ADDRESS (ES)
U.S. Army Research Office
P.O. Box 12211
Research Triangle Park, NC 27709-2211

10. SPONSOR/MONITOR'S ACRONYM(S)
ARO

11. SPONSOR/MONITOR'S REPORT NUMBER(S)
58523-LS-DRP.32

12. DISTRIBUTION AVAILIBILITY STATEMENT
Approved for Public Release; Distribution Unlimited

13. SUPPLEMENTARY NOTES
The views, opinions and/or findings contained in this report are those of the author(s) and should not contrued as an official Department of the Army position, policy or decision, unless so designated by other documentation.

14. ABSTRACT
This project has sought to better understand and predict outbreaks of infectious disease over time and space. It has endeavored to extend phylogenetic methods to understand the mutation and recombination events among pathogen genomes associated with the emergence of infectious diseases, and to layer this information with phenotypic and geographic data, enhanced with avian tissue collections. In this final phase of this project, we continued to focus on the use of large datasets of genetic sequences for hosts and pathogens, while working on phylogenetic visualization and also collecting avian specimens pertinent to the realm of biogeographical and host-pathogen research. We

15. SUBJECT TERMS
Influenza viruses, infectious diseases, genomic sequencing, phylogenetic analysis, Geographical Information Systems, avian host collections, science education

| 16. SECURITY CLASSIFICATION OF: | | | 17. LIMITATION OF ABSTRACT | 15. NUMBER OF PAGES | 19a. NAME OF RESPONSIBLE PERSON |
|---|---|---|---|---|---|
| a. REPORT | b. ABSTRACT | c. THIS PAGE | | | Ward Wheeler |
| UU | UU | UU | UU | | 19b. TELEPHONE NUMBER 212-769-5754 |

## Report Title

Continuation of Novel Analytical and Empircal Approaches to the Origin and Prediction of Pathogenicity

## ABSTRACT

This project has sought to better understand and predict outbreaks of infectious disease over time and space. It has endeavored to extend phylogenetic methods to understand the mutation and recombination events among pathogen genomes associated with the emergence of infectious diseases, and to layer this information with phenotypic and geographic data, enhanced with avian tissue collections. In this final phase of this project, we continued to focus on the use of large datasets of genetic sequences for hosts and pathogens, while working on phylogenetic visualization and also collecting avian specimens pertinent to the realm of biogeographical and host-pathogen research. We extended our research by generalizing our approach to phylogenetic networks, and helped to address DOD's and the nation's STEM future workforce needs by creating project-related science education programs and resources for 7th–12th grade science teachers and students.

**Enter List of papers submitted or published that acknowledge ARO support from the start of the project to the date of this printing. List the papers, including journal references, in the following categories:**

**(a) Papers published in peer-reviewed journals (N/A for none)**

| Received | | Paper |
|---|---|---|
| 04/10/2013 | 23.00 | Daniel A. Janies, Jonathon Studer, Samuel K. Handelman, Gregorio Linchangco. A comparison of supermatrix and supertree methods for multilocus phylogenetics using organismal datasets, Cladistics, (02 2013): 0. doi: 10.1111/cla.12014 |
| 06/25/2013 | 24.00 | Lili Zhuang, Noel Cressie, Laura Pomeroy, Daniel Janies. Multi-species SIR models from a dynamical Bayesian perspective, Theoretical Ecology, (06 2013): 0. doi: 10.1007/s12080-013-0180-x |
| 08/02/2012 | 10.00 | Daniel A. Janies, Laura W. Pomeroy, Jacob M. Aaronson, Samuel Handelman, Jori Hardman, Kevin Kawalec, Thomas Bitterman, Ward C. Wheeler. Analysis and visualization of H7 influenza using genomic, evolutionary and geographic information in a modular web service, Cladistics, (05 2012): 0. doi: 10.1111/j.1096-0031.2012.00401.x |
| 08/02/2012 | 12.00 | Ward C. Wheeler, John S.S. Denton. Indel information eliminates trivial sequence alignment in maximum likelihood phylogenetic analysis, Cladistics, (05 2012): 0. doi: 10.1111/j.1096-0031.2012.00402.x |
| 08/02/2012 | 11.00 | Lavanya Kannan, Ward C Wheeler. Maximum Parsimony on Phylogenetic Networks, Algorithms for Molecular Biology, (05 2012): 0. doi: 10.1186/1748-7188-7-9 |
| 08/12/2013 | 27.00 | Andrés Varón, Ward C Wheeler. The tree alignment problem, BMC Bioinformatics, (11 2012): 0. doi: 10.1186/1471-2105-13-293 |
| 08/12/2013 | 29.00 | Ward C. Wheeler. Maximum a posteriori probability assignment (MAP-A): an optimality criterion for phylogenetic trees via weighting and dynamic programming, Cladistics, (07 2013): 0. doi: 10.1111/cla.12046 |
| 08/12/2013 | 28.00 | Ward C Wheeler, Andrés Varón. Local search for the generalized tree alignment problem, BMC Bioinformatics, (02 2013): 0. doi: 10.1186/1471-2105-14-66 |
| 08/28/2012 | 14.00 | Sivananinthaperumal Balachandran, Taej Mundkur, David C. Douglas, John Y. Takekawa, Stephen G. Willis, Nichola J. Hill, Kyle A. Spragens, Daniel Janies, Igor O. Voronkin, Diann J. Prosser, Baoping Yan, Scott H. Newman, Fumin Lei, Nyambayar Batbayar, Tseveenmyadag Natsagdorj, Charles M. Bishop, Patrick J. Butler, Martin Wikelski. Eco-Virological Approach for Assessing the Role of Wild Birds in the Spread of Avian Influenza H5N1 along the Central Asian Flyway, PLoS ONE, (02 2012): 0. doi: 10.1371/journal.pone.0030636 |
| 09/19/2011 | 1.00 | Daniel Janies, P.J. Embi, P.R. Payne. Health-care hit or miss?: Collect genetic data on pathogens, Nature, (02 2011): 327. doi: |
| 09/19/2011 | 2.00 | S. Bokhari, L. Pomeroy, D. Janies. Reassortment Networks and the Evolution of Pandemic H1N1 Swine-origin Influenza, IEEE Transactions on Computational biology and bioinformatics, (04 2011): 0. doi: |
| 09/19/2011 | 3.00 | S. Kumar, S. Handelman, I. Voronkin, V. Mwapasa, D. Janies, S. Rogerson, S. Meshnick, J. Kwiek. Different Regions of HIV-1 Subtype C Env are Associated with Placental Localization and in Utero Mother-to-Child Transmission. , Journal of Virology, (05 2011): 0. doi: |

11/21/2013 26.00 Surender Kumar, Jose Bazan, Megan Mefford, Jessica Mates, Igor Voronkin, Samuel Handelman, Victor Mwapasa, William Ackerman IV, Daniel Janies, Jesse Kwiek. Genotypic and phenotypic heterogeneity in the U3R region of HIV-1 subtype C ,
AIDS Research and Human Retroviruses, (07 2013): 0. doi:

**TOTAL:** **13**

**Number of Papers published in peer-reviewed journals:**

## (b) Papers published in non-peer-reviewed journals (N/A for none)

Received Paper

**TOTAL:**

**Number of Papers published in non peer-reviewed journals:**

## (c) Presentations

**Number of Presentations:** 0.00

## Non Peer-Reviewed Conference Proceeding publications (other than abstracts):

Received Paper

**TOTAL:**

**Number of Non Peer-Reviewed Conference Proceeding publications (other than abstracts):**

## Peer-Reviewed Conference Proceeding publications (other than abstracts):

Received Paper

**TOTAL:**

**Number of Peer-Reviewed Conference Proceeding publications (other than abstracts):**

---

## (d) Manuscripts

| Received | | Paper |
|---|---|---|
| 01/06/2012 | 6.00 | Lavanya Kannan, Ward C. Wheeler. Maximum Parsimony on Phylogenetic Networks, Algorithms in molecular biology (01 2012) |
| 01/11/2013 | 17.00 | Ward C. Wheeler. MAXIMUM A POSTERIORI PROBABILITY ASSIGNMENT(MAP-A): AN OPTIMALITY CRITERION FOR PHYLOGENETICTREES VIA WEIGHTING AND DYNAMIC PROGRAMMING, SIAM Journal on Scientific Computing (01 2013) |
| 01/17/2013 | 18.00 | Lili Zhuang, Noel Cressie, Laura Pomeroy, Danial Janies. Multi-species SIR Models from a Dynamical Bayesian Perspective, Theoretical Ecology (01 2013) |
| 01/28/2014 | 30.00 | Lavanya Kannan, Ward Wheeler. On exactly computing the parsimony scores on phylogenetic networks using dynamic programming, Journal of Computational Biology (10 2013) |
| 01/28/2014 | 31.00 | Laura Pomeroy, Chris Krueger, Igor Voronkin, Jori Hardman, Yuqi Zhang, Rasmus Hovmöller, Daniel Janies. Evolution and spread of H7 influenza A viruses over time, space, and various hosts, Transboundary and Emerging Diseases (01 2014) |
| 01/30/2012 | 7.00 | John S. S. Denton, Ward C. Wheeler. Indel information eliminates trivial sequence alignment in maximum likelihood phylogenetic analysis, Cladistics (01 2012) |
| 02/01/2013 | 19.00 | Jessica Mates, Surender Kumar, Jose Bazan, Igor Voronkin, Samuel Handelman, Victor Mwapasa, William Ackerman IV, Daniel Janies, Jesse J Kwiek. Genotypic and phenotypic heterogeneity in the U3R region of HIV-1 subtype C, AIDS Research and Human Retroviruses (01 2013) |
| 04/11/2013 | 22.00 | Ward C. Wheeler. Maximum a Posteriori Probability Assignment (MAP-A): An Optimality Criterion for PhylogeneticTrees via Weighting and Dynamic Programming, Cladistics (02 2013) |
| 04/11/2013 | 21.00 | Ward C. Wheeler, Peter M. Whitely. Words as Sequences: Uto-Aztecan Language Evolution and Biogeography, Cladistics (03 2013) |
| 04/11/2013 | 20.00 | Ward C. Wheeler. Phyletic Groups on Networks, Cladistics (03 2013) |
| 06/25/2013 | 25.00 | Samuel Handelman, Jacob Aaronson, Igor Voronkin, Jesse Kwiek, Joseph Verducci, Daniel Janies. Cladograms with Path to Event for Viruses (ClaPTE/V): A novel algorithm to detect genotype orphenotype associations on phylogenies., BMC Bioinformatics (06 2013) |
| 07/10/2012 | 8.00 | Andrés Varón, Ward C Wheeler. The Tree Alignment Problem, BMC Bioinformatics (06 2012) |
| 07/10/2012 | 9.00 | Andrés Varón, Ward C Wheeler. Local Search for the Generalized Tree Alignment, BMC Bioinformatics (06 2012) |

08/31/2012 15.00 Samuel Kenneth Handelman, Ph.D., Jacob M. Aaronsom, Jesse J. Kwiek, Joseph S. Verducci, Daniel Janies. Phylogeny with Path to Event for Viruses (PhyloPTE/V): Robust efficient associationdetection for qualitative traits on phylogenies.,
Public Library of Science: Computational Biology (08 2012)

12/29/2011 5.00 Scott Newman, Nichola Hill, Kyle Spragens, Daniel Janies, Igor Voronkin, Diann Prosser, Yan Baoping, Fumin Lei, Nyambayar Batbayar, Natsagdorjin Tseveenmyadag, Charles Bishop, Patrick Butler, Martin Wikelski, Balachandran Sivananinthaperumal, Taej Mundkur, David Douglas, John Y. Takekawa. An Eco-Virological Approach for Assessing the Role of Wild Birds in the Spread of Avian Influenza H5N1 Along the Central Asian Flyway,
PLoS ONE (12 2011)

12/29/2011 4.00 Daniel Janies, Laura Pomeroy, Jacob Aaronson, Samuel Handelman, Jori Hardman, Kevin Kawalec, Thomas Bitterman, Ward Wheeler. Analysis and visualization of H7 influenza using genomic, evolutionary and geographic information in a modular web service,
Cladistics (12 2011)

12/31/2012 16.00 Daniel Janies, Jonathan Studer, Samuel Handelman, Gregorio Linchangco. A comparison of supermatrix and supertree methods for multilocus phylogenetics using organismal datasets,
Cladistics (12 2012)

**TOTAL:** **17**

**Number of Manuscripts:**

## Books

Received Paper

08/28/2012 13.00 Ward C. Wheeler. Systematics, Chichester, UK: John Wiley & Sons, Ltd, (04 2012)

**TOTAL:** **1**

## Patents Submitted

## Patents Awarded

## Awards

## Graduate Students

| NAME | PERCENT_SUPPORTED |
|---|---|

**FTE Equivalent:**
**Total Number:**

## Names of Post Doctorates

| NAME | PERCENT_SUPPORTED |
|---|---|
| Lavanya Kannan | 1.00 |
| Louise Crowley | 1.00 |
| Chris Krueger | 1.00 |
| **FTE Equivalent:** | **3.00** |
| **Total Number:** | **3** |

## Names of Faculty Supported

| NAME | PERCENT_SUPPORTED | National Academy Member |
|---|---|---|
| Daniel Janies | 0.06 | |
| **FTE Equivalent:** | **0.06** | |
| **Total Number:** | **1** | |

## Names of Under Graduate students supported

| NAME | PERCENT_SUPPORTED |
|---|---|
| **FTE Equivalent:** | |
| **Total Number:** | |

## Student Metrics

This section only applies to graduating undergraduates supported by this agreement in this reporting period

The number of undergraduates funded by this agreement who graduated during this period: ..... 0.00

The number of undergraduates funded by this agreement who graduated during this period with a degree in science, mathematics, engineering, or technology fields:..... 0.00

The number of undergraduates funded by your agreement who graduated during this period and will continue to pursue a graduate or Ph.D. degree in science, mathematics, engineering, or technology fields:..... 0.00

Number of graduating undergraduates who achieved a 3.5 GPA to 4.0 (4.0 max scale):..... 0.00

Number of graduating undergraduates funded by a DoD funded Center of Excellence grant for Education, Research and Engineering:..... 0.00

The number of undergraduates funded by your agreement who graduated during this period and intend to work for the Department of Defense ..... 0.00

The number of undergraduates funded by your agreement who graduated during this period and will receive scholarships or fellowships for further studies in science, mathematics, engineering or technology fields: ..... 0.00

## Names of Personnel receiving masters degrees

| NAME |
|---|
| **Total Number:** |

## Names of personnel receiving PHDs

| NAME |
|---|
| **Total Number:** |

## Names of other research staff

| NAME | PERCENT_SUPPORTED |
|---|---|
| Nicholas Lucaroni | 1.00 |
| **FTE Equivalent:** | **1.00** |
| **Total Number:** | **1** |

## Sub Contractors (DD882)

## Inventions (DD882)

## Scientific Progress

See Attachment.

## Technology Transfer

**CONTINUATION OF NOVEL ANALYTICAL AND EMPIRICAL APPROACHES TO THE ORIGIN AND PREDICTION OF PATHOGENICITY**

ARO Grant #W911NF-10-1-0339

*Final Report*

January 31, 2014

## TABLE OF CONTENTS

## LIST OF APPENDICES, ILLUSTRATIONS, AND TABLES

**CONTINUATION OF NOVEL ANALYTICAL AND EMPIRICAL APPROACHES TO THE ORIGIN AND PREDICTION OF PATHOGENICITY**

ARO Grant #W911NF-10-1-0339

*Final Report*

January 31, 2014

We are pleased to submit the following final report for the project, *Continuation of Novel Analytical and Empirical Approaches to the Origin and Prediction of Pathogenicity* (#W911NF-10-1-0339), which began at the American Museum of Natural History (AMNH) on September 1, 2010 and ended October 31, 2013.

## OVERVIEW

This project has sought to better understand and predict outbreaks of infectious disease over time and space. It has endeavored to extend phylogenetic methods to understand the mutation and recombination events among pathogen genomes associated with the emergence of infectious diseases, and to layer this information with phenotypic and geographic data, enhanced with avian tissue collections. In this final phase of this project, we continued to focus on the use of large datasets of genetic sequences for hosts and pathogens, while working on phylogenetic visualization and also collecting avian specimens pertinent to the realm of biogeographical and host-pathogen research. We extended our research by generalizing our approach to phylogenetic networks, and helped to address DOD's and the nation's STEM future workforce needs by creating project-related science education programs and resources for $7^{th}$–$12^{th}$ grade science teachers and students.

We are pleased to report that the proposed project deliverables were successfully met. Major accomplishments in critical task areas are summarized below, followed by detailed discussion of activities and achievements in each area.

## SUMMARY OF CRITICAL TASKS AND IMPORTANT RESULTS

- Theoretical and Algorithmic Research and Software Development
  We created a new model for phylogenetic analysis of genomic data from host-pathogen complexes. We completed implementation of statistical phylogenetic models, as well as theoretical work on phylogenetic networks while designing new algorithms and implemented them into a new release of phylogenetic software, POY5.1.1. Through the development of flexible, open, and robust theories and tools, the project has expanded the ability of scientists to test hypotheses of viral (and other organismal) evolution.
- Field and Collections-based Research on the Evolution of Avian Hosts
  We significantly enhanced the sampling of species capable of dispersing avian influenza, collected avian tissues from targeted regions around the world, and deposited the samples into AMNH's Ambrose Monell Cryo Collection (the Monell Collection), making the genetic materials widely available to researchers. We also generated large-scale phylogenetic and migratory maps for thousands of avian species, providing an important template for studying host-pathogen co-evolution.

- Development of Visualization Techniques
  We developed a means to integrate project data into user-friendly geographic information system (GIS) tools and communicated our research to a wide audience of public health stakeholders, with published peer-reviewed papers including many in infectious disease.[1] We extended our web-based application, SUPRAMAP (which uses POY to produce files that Google Earth reads), to allow for tracking of host-pathogen complexes across time and space, and made improvements to the graphical interface for visualization of character evolution. By developing code that combines virtual globe and bioinformatics technologies to perform hypothesis driven research on pathogens, we have demonstrated that these technologies can be combined in a meaningful way to address significant infectious disease problems.
- Outreach
  We have carried out active and diversified outreach to public health and other user communities in government and academia and to the public at home and abroad. The easy-to-use web-based applications that we developed served to increase widespread usage of our tools and resources in the different stakeholder communities.
- Development of STEM Education Activities
  Collaborating teams of AMNH project scientists and science educators successfully adapted and integrated project-related content and research into STEM education programs and nationally distributed science media, helping to strengthen the STEM pipeline. We delivered:
    - Professional development programs with multimedia resources for grades 7–12 science teachers; a teaching case used in professional development and available online; and an eight-minute documentary video for the Museum's nationally distributed science media program on the evolution of pathogenicity in *Staphylococcus aureus* (MRSA).
    - Out-of-school college readiness programming and mentoring for high school students, incorporating the research and work of project scientists.

**Theoretical and Algorithmic Research and Software Development**

Seeking to advance understanding and tools for examining the evolution of infectious diseases, our goal in this project component was to generalize the theory and tools to use host-pathogen complexes as data points that undergo the addition, loss, or transfer of genetic material. We have added to the existing POY code base the ability to analyze host-pathogen complex scenarios more completely and realistic hypotheses for infection, co-infection, and host shifts, and have completed the analysis of these networks.

We have been able to add this functionality to our code base through new algorithmic advances (Kannan and Wheeler, 2013, in press; Wheeler, in press; Varón and Wheeler,

[1] For example, infectious diseases and mathematics (Vinh et al., 2007; Janies and Pol, 2008), H7 influenza (Bulach et al., 2010; Janies et al., 2012); H5 influenza (Janies et al., 2007; 2011; Hill et al., 2009; Hovmöller, 2010; Bokhari and Janies, 2010; Newman et al., 2012); H1 influenza (Janies et al., 2010a,b; Bokhari et al., 2012; Zhang et al., 2013); HIV (Kumar et al., 2011; Mates et al., 2013); SARS-CoV and other Coronaviruses (Zhang et al., 2007a,b; Janies et al; 2008).

2012, 2013) and implementation (POY, version 5.1.1 released; Wheeler et al., in prep). All of these advances are now available to researchers. We have also modularized our code (PhylOcaml, https://github.com/amnh/phylocaml) such that other investigators can more easily import and use the codes we have developed.

Over the course of this project, there have been four main areas of algorithmic and implementation progress. These are in statistical (maximum likelihood) procedures for phylogenetic analysis of unaligned sequences, Bayesian phylogenetic approaches, network graph optimization, and host-pathogen interaction phylogenetics.

Our first task was to create statistical likelihood-based phylogenetic optimization procedures that could be applied to unaligned sequence data. This required the application of alternate forms of maximum-likelihood ("most-parsimonious likelihood") due to the complexity of the tree-alignment problem in a model-based context. We first completed a beta implementation of our statistical phylogenetic methods and released our software implementation (POY5.0-beta) to the scientific community. After much testing and refinement, we released POY5.0 in August 2013. This release included our new facilities for identifying multiple models for different sequence data partitions.

Subsequent to the likelihood work, we created a Bayesian analysis procedure based on maximum posterior probability sequence assignments (MAP-A; Wheeler, 2013). Completed implementation of Bayesian statistical framework followed the theoretical work and this approach has been extended to networks (Wheeler, 2013). Due to the additional time complexity burden of these statistical methods, we have completed the implementation of a second means of parallelization at the level of refinement neighborhood.

With statistical models in hand, we developed polynomial-time algorithms to identify optimal solutions for sequence changes on networks. These allow not only vertical transmission of information but also hybridization, horizontal exchange, and reassortment of lineages. Our initial implementation compared a brute-force solution of all possible sequence assignments to a polynomial-time dynamic programming-based approach. We have seen a large improvement in execution time with no loss of optimality. Our initial network code approach (Kannan and Wheeler, 2012; Kannan and Wheeler, in press) and model have also been applied to human language evolution (Wheeler, 2013b; Wheeler and Whiteley, submitted).

We have completed as well the initial algorithmic and prototype work on our second iteration of network optimization code and are testing this approach. A manuscript describing this work is in press.

In sum, with DARPA support we created a new model for phylogenetic analysis of genomic data from host-pathogen complexes. These complexes may be pathogens and hosts alone, or in combination as infected hosts. We have developed a theoretical framework to evaluate complex tree histories involving co-evolution and host shifts. Algorithms have been designed and implemented into our existing base of code (POY4; Varón et al., 2010; POY5). Based on this work, we applied prototype network codes to

biogeographic and microbial data and integrated them into POY code base and released POY5.1 with new parallel codes in December 2013. Source, binaries, and documentation are available at http://www.amnh.org/our-research/computational-sciences/research/projects/systematic-biology/poy.

We have met our goals for integrating of statistical models to unaligned sequence-based trees and networks and applying them to human pathogens. The algorithmic and implementation work we have carried out enables us and others to study the origin of evolutionary novelties in pathogenic and non-pathogenic lineages (e.g., viral strains) using the broad variety of mathematical approaches that exist in phylogenetic research today.

**Field and Collections-based Research on the Evolution of Avian Hosts**

Providing an important collections-based link to the other project elements, this part of the project focused on advancing understanding of the spread of diseases via birds and their migratory behavior. We enhanced sampling of species capable of dispersing avian influenza into new communities and ecosystems; preserved the samples in the AMNH's Monell Collection, which serves as an important and unique resource for researchers working to understand the evolution and biogeography of avian hosts of influenza, and generated critical baseline data on the spread of diseases via birds. Using databases of genomic and migratory data, we generated large-scale phylogenetic and migratory maps for thousands of avian species, providing an important template for the study of host-pathogen co-evolution.

To enhance sampling of avian species capable of dispersing avian influenza, we conducted very successful collecting trips for tissue samples from biologically distinct regions that have taxa that are underrepresented in avian tissue collections and obtained additional samples through several salvage operations.

In 2011, we carried out a three-week field trip to Western Australia, a wintering area for many East Palaearctic breeders, having obtained a collecting permit from the Western Australian Department of Environment and Conservation (WADEC) for up to five specimens per species of a wide variety of birds. We worked in forest, woodland, shrub, and heath habitats, and traveled to three remote collecting sites within the region in order to maximize taxonomic coverage. Birds were sampled in two ways: field collecting by mist-nets and from frozen salvaged specimens gathered from museums, the WADEC, and rehabilitation centers. We saved anatomical and tissue samples from all specimens.

On this trip, 182 birds of 80 species were sampled, including many endemics. Of these, 55 species were new additions to AMNH's Monell Collection. In addition, we preserved important, data-rich, anatomical voucher specimens which were lacking in the AMNH collections.

In 2002, we had another successful field period of a four-week duration. The primary field work was carried out in the northern part of the Kimberley Region of Western Australia. This area is recognized as a globally important biodiversity hotspot, and the

bird fauna contains many endemic taxa. The region is also the wintering grounds for many species of bird that breed in northeast Asia. General collecting permits were obtained by our collaborators at the Western Australian Museum in Perth.

Over two weeks in May, we travelled across the Kimberley via the Gibb River Road between Broome and Kununurra, an unpaved road only accessible during the dry season. Three collecting camps were established along this road. In total, we collected 280 specimens of 60 species in this region.

After completing the work in the north, we spent five days collecting in heathland habitat on the south coast in the Fitzgerald River National Park. For this phase of the expedition, we obtained permits from WADEC for up to five specimens per species of a wide variety of birds. These specimens supplemented our collections from the southwest that were collected on a previous expedition in November 2010. Because this area is temperate, the composition of the bird fauna varies between winter and summer. We collected a total of 75 specimens of 26 species in this area.

In addition to field work, we were able to obtain salvaged birds from the Western Australian Museum, the Broome Bird Observatory, and the Kununurra Department of Environment and Conservation. This added an additional 60 samples of 44 species.

During the four-week field period, a total of 415 birds of 113 species were sampled for tissue samples and anatomical specimens. Included in this sample were many of the regional endemics. Of the samples collected, 49 species were new additions to the Monell Collection. In addition, we preserved many important, data-rich, anatomical voucher specimens which were lacking in the AMNH collections.

The tissues obtained during this fieldwork significantly enhanced our sampling of species capable of dispersing avian influenza, with the genetic material now widely available to researchers through the Monell Collection. Importantly, DNA sequences obtained from these tissues were incorporated into genomic databases and, by combining these datasets with phenotypic and migratory data, we generated large-scale phylogenetic and migratory maps for thousands of avian species that provided a template for studying host-pathogen co-evolution.

## Development of Visualization Techniques

In this project component, we used several means of visualization and service implementations. Our goals were to join infrastructure, communication, and hypothesis testing into a common workflow. This approach allowed many professionals with diverse skill sets to work together in a team science approach. Our work represents the first instance of applying phylogenetic character evolution techniques in a GIS. A use case in infectious diseases allowed scientists to not only put strains in a GIS but also the properties of strains as encoded by key mutations (e.g., E627K in PB2 in H5N1 or H7N9, which confers increased replication in mammals) (Janies et al., 2007; Pomeroy et al., submitted).

Users of our tools are able to build phylogenetic maps of lineages of avian influenza, which provides a template for mapping and investigating host-pathogen coevolution (e.g., Newman et al., 2012) and the molecular evolution of pathogenicity across geography (Janies et al., 2012). In our work on H7 influenza we found that the emergence of pathogenic strains of H7 is highly labile. Many transitions from high pathogenicity to low pathogenicity and from low pathogenicity to high pathogenicity occur. In 2012, we identified several lineages of H7 influenza with biomarkers of high pathogenicity circulating regions that had not been reported in the scientific literature.

We have used SUPRAMAP (http://supramap.org), a web-based application for studying biogeographic, genotypic, and phenotypic evolution, to study drug resistance and host shifts in influenza and host shifts in coronaviruses across time and space. The original implementation of SUPRAMAP was built with tightly coupled client and server software (Janies et al., 2010). We decoupled the components to provide a modular processing service (POYWS; poyws.org. Figure 1) that can be consumed by a data provider or application developer to build a novel tool (Janies et al., 2012).

To illustrate the service we also produced a novel application, GEOGENES (formerly GISBANK, which we renamed in 2012 due to the overlap of GISBANK with other projects). Unlike SUPRAMAP, in which the user is required to create and upload data files, in GEOGENES the user works from a graphical interface to query an underlying dataset. The processing service can be coupled with diverse data sources supporting a variety of applications across the natural and biomedical sciences.

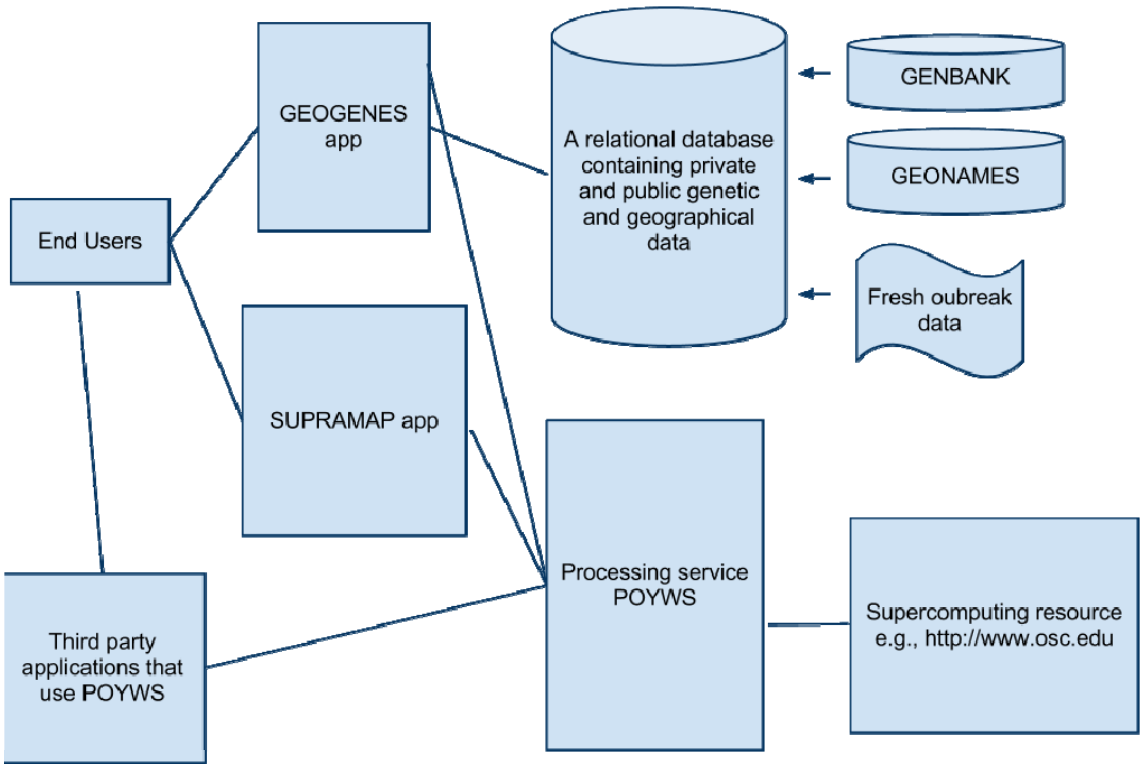



**Figure 1. Diagram of the interactions of the user, applications (e.g., SUPRAMAP, GEOGENES), internal (e.g., fresh outbreak data) and external data sources (e.g., GEONAMES, GENBANK), and computing resources (POYWS and OSC).**

GEOGENES is hosted on a shared, scheduled system, the Glenn cluster at the Ohio Supercomputer Center (OSC; https://geogenes.osc.edu/menu/index). It is a simple and flexible web service for creating applications for phylogenetic analyses and visualization that is useful across a broad variety of scientific investigations. It increases the usability of our software by providing an entirely graphically driven query interface to the underlying data and processing services. The graphical interface we have created allows a user to select data from a variety of sources (e.g., various genes, taxa, hosts, times, geographic regions) via a query and frees the users from editing input files, a process which is error-prone. After a query is processed, GEOGENES allows the user to submit the data for execution in the processing service. Within an individual user account, the queries and results can be stored for later processing, for example, with updated data. Queries and results can be shared with other users such that the workflow underlying the science is repeatable.

Many of the aspects of comparative genomic research, such as multiple sequence alignment and tree search, require large numbers of processors only available in high performance computers (HPC; e.g., a scheduled supercomputer, dedicated cluster, multicore server, grid, or cloud). Often, the set-up required of the user for high

performance systems presents a significant hurdle. In order to remove the burdens of acquisition and maintenance of hardware, installation of languages and middleware, compilation of applications, configuration of accounts, production of scripts, navigation of schedulers, and work with a command line interface, we have developed a processing service that executes these analyses on many central processing units via the click of a button on a webpage.

The first version of SUPRAMAP ran on a small computing cluster, and its client and server were tightly coupled (Janies et al., 2010a). The generic and modular web service that we have now created runs POY (Varón et al., 2010) with SUPRAMAP as a plug-in under a scheduler on a large cluster. In addition to GEOGENES, SUPRAMAP currently consumes the processing service.

The SUPRAMAP web-based application was installed with a new host Renaissance Computing Institute, RENCI, Chapel Hill, NC in 2013. In the current version of SUPRAMAP, we have implemented the ability for users to trace phenotypic characters. Previously this function was only available via scripts. We improved the interface for visualization of character evolution. Now users do not have to use scripts to color trees to trace key mutations or phenotypes (Figure 2). These functionalities are provided in the graphical user interface as another optional step in the workflow ("color KML").

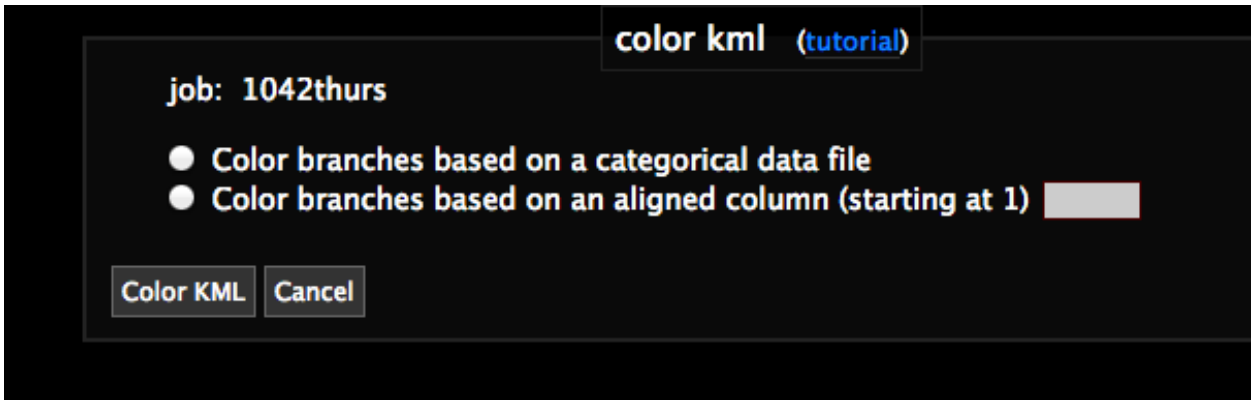



**Figure 2. New graphical interface for coloring KML.**

In the area of testing hypotheses on identifying exchanges of pathogens and pathogen genes among hosts in time and space, we have devoted our time to the recent outbreak of H7N9 in China. We have created SUPRAMAPs and ROUTEMAPs describing the host and geographic origins and movement of each segment. Here is a summary of our results:

- The China-Taiwan H7N9 outbreak in early 2013 was caused by a reassortant virus made primarily of genetic segments previously circulating in chickens in China. However, some genetic segments have anseriform host ancestry and some segments have international connections to and from China-Taiwan to neighboring countries.

- For H7N9-2013, we have evidence for multiple independent post-outbreak transmissions to anseriformes, columbiformes, and passeriformes. These results

indicate that, despite the abatement of the recent outbreak, the H7N9 virus may spread beyond China-Taiwan as has been seen in H5N1.

This work by Pomeroy et al. has been submitted to *Transboundary and Emerging Diseases* for review.

**Outreach**

Outreach has been an important and successful project component. Our outreach efforts during this final phase focused on teaching and gathering feedback about our developed tools, SUPRAMAP and POY software, and sharing project information. In the first project year, we focused outreach on the opportunity to reach a wide range of cross-disciplinary scientists by co-convening the conference, Next Generation Sequencing: Transformative Technology for Biodiversity Science, held in Washington, DC on April 18–19, 2011. AMNH co-convened this conference with the FDA, the Smithsonian Institution's National Museum of Natural History, and the George Washington University. The meeting brought together a broad and cross-disciplinary group of algorithmic and bioinformatics researchers from academia, research institutions, and government concerned with biodefense, food safety, and related security issues. DOD, Department of Energy, FDA, USDA, and USFWS were among the federal agencies represented. Principal Investigator Dr. Ward Wheeler presented on whole genome phylogenetics with POY, and UNC Charlotte Project Director Dr. Daniel Janies presented on SUPRAMAP.

Throughout the project, we have actively disseminated project information and results in a variety of venues and media. In addition to software updates, we produced a number of peer-reviewed publications, email and user group support online, visits, telconferences, and demonstrations with groups including the Global Initiative on Sharing Avian Influenza Data (GISAID), National Institute of Allergy and Infectious Diseases (NIAID), and the Influenza Research Database (FLUDB.org).

Public outreach has also been achieved via the Internet (e.g.:
http://www.amnh.org/explore/news-blogs/research-posts/researchers-broaden-reach-of-virus-tracking-software;
http://www.eurekalert.org/pub_releases/2012-05/osc-rtv052212.php;
http://www.sciencenewsline.com/technology/2012052306400034.html;
http://www.hpcwire.com/hpcwire/2012-05-23/bioinformatics_researcher_software_engineers_look_to_improve_genomics_software.html;
http://geospatialworld.net/News/View.aspx?id=24842_Article;
http://apb.directionsmag.com/entry/dr.-oz-likes-gis-and-other-health-gis-news/253955;
http://www.cidrap.umn.edu/cidrap/content/influenza/general/news/may2312newsscan.html).

In addition AMNH held a public program (SciCafe) for general audiences featuring Dr. Janies on February 2, 2012, and the research was featured in a radio interview (http://wosu.org/2012/news/2012/05/24/osu-researcher-touts-new-epidemic-tracking-

software/). We are very pleased that SUPRAMAP images appear in the National Museum of Natural History's current exhibition *Genome: Unlocking Life's Code* (http://www.mnh.si.edu/exhibits/genome/).

**Development of STEM Education Activities**

A STEM education component was carried out in the project's last period. It included the creation of project-related programs and resources for science teachers and students in grades 7–12, as we adapted content, themes, and questions driving the project research into materials suited for two education areas: (1) Teacher Professional Development and (2) Out-of-School College Readiness Programming for High School Students.

- Teacher Professional Development

Three two-day professional development workshops planned for secondary science teachers (aimed at New York City high school Living Environment teachers) were offered on July 16 to 17, 2013, July 30 and August 2, 2013, and August 15 to 16, 2013 with 31, 35, and 30 teachers attending, respectively. An additional workshop, funded by other AMNH sources, will be offered in February 2014; as of the date of this report, 20 teachers had already enrolled.

We identified the following high school learning outcome from the Next Generation Science Standards around which to focus our efforts: obtain, evaluate, and communicate information describing how changes in environmental conditions can affect the distribution of traits in a population causing 1) increases in the population of some species, 2) the emergence over time of new species, and 3) the extinction of other species. This outcome framed the learning experiences we developed for participants using the teaching case resources described below, and informed the development of a related activity using the Cold Spring Harbor DNA subway interactive website (http://dnasubway.iplantcollaborative.org) to support an inquiry-based activity designed for students using DNA sequences. The participating teachers also engaged in two writing exercises, one aligned to the Common Core Literacy Standards and one to the Next Generation Science Standards, with emphasis on explanatory writing and obtaining, evaluating, and communicating information.

We also produced an eight-minute documentary video for AMNH's *Science Bulletins*—a nationally distributed multimedia program that makes breaking science news and current research accessible to the general public—that was further enhances, as described below for professional development uses. Centering on the evolution of pathogenicity in *Staphylococcus aureus* (MRSA), the video features the work of two scientists: Dr. Paul Planet at Columbia University, who has identified and dated a gene transfer event that enhanced MRSA's ability to colonize on skin and cause disease; and Dr. James Meadows, of the Center for Biology in the Built Environment at University of Oregon, whose work focuses on the origins of pathogenesis in the context of the human microbiome. It explores how genes that certain MRSAs have taken up have allowed them to be more infectious/successful in this "environment" of the human body, as well as how scientists are beginning to approach the dispersal of bacteria through the environments

people live in—a critical line of inquiry now that MRSA has moved from the hospital into the community. This *Science Bulletins* feature for general audiences debuted in the Museum's Hall of Human Origins in September 2013 and is also available online (http://www.amnh.org/explore/science-bulletins/(watch)/human/documentaries/when-good-bacteria-go-bad/(p)/1) and to our subscribers, which include museums, science centers, and universities across the nation that feature *Science Bulletins* at their locations.

To utilize this content for professional development, AMNH completed production of an associated four-segment video for the teaching case that further explicates the scientific process and discoveries communicated by the MRSA documentary as well as providing accompanying written passages for both teachers and students. The student versions were finalized after receiving feedback in the summer 2013 professional development workshops, and in November 2013 all of these materials were made available on the AMNH website (http://www.amnh.org/explore/curriculum-collections/bacteria-evolving-tracing-the-origins-of-a-mrsa-epidemic).

We are integrating these new resources into existing AMNH professional development programs to extend the educational benefits beyond the grant period, and, as mentioned, have already planned an additional two-day professional development workshop for February 19 to 20, 2014. Moreover, the teaching case will be incorporated into the AMNH's Teacher Renewal for Urban Science Teachers (TRUST)—an annual two-week summer professional development institute designed specifically to meet New York City's extensive need for well-prepared science teachers. In 2014, the TRUST institute in life science will focus on evolution and biodiversity.

An additional *Science Bulletin* video was developed with the collaborative assistance of Dr. Dan Janies, project director at University of North Carolina at Charlotte, who advised AMNH Education staff on *SUPRAMAP Tracks Diseases as They Evolve.* The video features SUPRAMAP in a two-minute data-based graphic animation, and as with all *Science Bulletins* media, it appears in the Museum's halls, at partner venues, and on the web (http://www.amnh.org/explore/science-bulletins/(watch)/human/news/supramap-tracks-diseases-as-they-evolve).

- Out-of-School College Readiness Programming for High School Students

DARPA funding served to enhance AMNH's existing educational offerings by leveraging the project's research for adaptation and incorporation into out-of-school courses for New York City students. Two sessions of the course entitled "Evolution of Viruses," which incorporated case studies and data from DARPA-funded postdoctoral scientist Ronald Clouse's research, were completed by 38 New York City high school students. Additionally, two sessions of the more advanced research-level "Comparative Genomics" course, which Dr. Clouse helped to modify and co-teach, were completed by 39 New York City high school students. Furthermore, two high school students completed a year-long mentored research project with Dr. Clouse.

The out-of-school course "Evolution of Viruses" was developed by Dr. Clouse and Nuala

Caomhanach, an educator with a strong background in the biological sciences who also taught the course. A fall session ran from September 10 through October 29, 2012, and a spring session ran from February 25 to April 19, 2013. Over 12 classes, students explored the following questions:

- Where do deadly viruses come from, and can we predict future outbreaks?
- Does virus DNA contain clues to their history, and can we reconstruct their paths to human hosts?

Using SARS as a case study, students learned the basics of viral evolution, the methods used by scientists to study it, and how this information can be used to prevent the next outbreak of disease.

The more advanced "Comparative Genomics" after-school course consisted of 11 classes and was offered to high school students twice—one session running from September 10 to October 29, 2012, and the other from April 23 to May 30, 2013. This course introduced students to the structure and function of the genome, techniques in molecular biology used in Museum laboratories, and the generation, transmission, and study of genetic variation in the context of evolution and species identification.

In June 23, 2013, students who participated in these courses joined us for a field trip to Cold Spring Harbor DNA Learning Center. They received a tour of the facilities and participated in a human genetic diversity and evolution lab.

In addition to these accomplishments, two high school students completed research mentorships with project scientists, as part of their participation in the AMNH's Science Research Mentoring Program, an intensive program that offers high school students the opportunity to join ongoing research projects led by AMNH scientists who provide them with mentored support. During the 2012–2013 school year, the students worked with postdoctoral scientists Clouse and Lavanya Kannan on their research involving novel phylogenetic analysis. The students prepared sequence data sets for a novel phylogenetic network analysis and successfully completed their research project, titled "Developing Novel Algorithms for Evaluating Reticulation Events." They presented a poster and an oral presentation at the Science Research Mentorship Program graduation on June 6, 2013.

**CONCLUSION**

With DARPA support, this project has made important contributions to knowledge about the evolution and prediction of infectious disease, notably enlarging our understanding of how pathogens originate, evolve, and spread over time and place. We have developed important new resources and tools for research and public health communities, with applications across the natural and biomedical sciences. We have enhanced sampling of and access to critical avian species and also created related educational programs and resources that support the DOD's and AMNH's shared commitment to building a strong STEM-capable workforce.

Appendix 1: Papers, Software, and Presentations
Covering the period of August 1, 2013 to October 31, 2013

***Manuscripts Submitted (but not yet published)***

Kannan, L., and Wheeler, W. C. On exactly computing the parsimony scores on phylogenetic networks using dynamic programming. *Journal of Computational Biology*. In press.

Pomeroy, L., Krueger, C., Voronkin, I., Hardman, J., Zhang, Y., Hovmöller, R., Janies, D. Evolution and spread of H7 influenza A viruses over time, space, and various hosts. *Transboundary and Emerging Diseases*.

***Papers***

Mates, J., Kumar, S., Bazan, J., Mefford, M., Voronkin, I., Handelman, S., Mwapasa, V., Ackerman, W., Janies, D., Kwiek, J. 2013. Genotypic and phenotypic heterogeneity in the U3R region of HIV-1 subtype C. *AIDS Research and Human Retroviruses*. http://online.liebertpub.com/doi/abs/10.1089/AID.2013.0026?mi=3yvu6q&af=R&searchText=breastfeeding&target=default&

Zhang, L., Cressie, N., Pomeroy, L., Janies, D. 2013. Multi-species SIR Models from a Dynamical Bayesian Perspective. *Theoretical Ecology*. http://link.springer.com/article/10.1007%2Fs12080-013-0180-x

***Software and Web Applications***

2011-present — POY WEBSERVICE http://poyws.org
Simple and flexible web service for creating applications for sequence alignment, phylogenetic analyses, and geographic visualization. Useful across a broad variety of scientific fields including public health, agricultural, and natural sciences. Published as Janies et al., 2012.

2011-present — GEOGENES http://geogenes.org
A web-based application that consumes POYWEBSERVICE but, in contrast to SUPRAMAP, provides a graphical interface to query underlying datasets of genetic and geographic data for pathogens. Published as Janies et al., 2012.

2009-present — ROUTEMAP http://routemap.osu.edu
A web-based application to analyze disease transmission events implied by genetic sequence data from pathogens. Published as Hovmöller et al., 2010.

2007-present — SUPRAMAP http://supramap.org
A web-based application to produce phylogenetic trees and GIS layers of the spread of pathogens across time space and various hosts. Published as Janies et al., 2010; 2012. Code for the SUPRAMAP plug-in is public and part of the repository for POY. http://code.google.com/p/poy/

***Presentations***

Janies — October 4, 2013. “Large dataset analysis as applied to genomics and geography of infectious diseases.” College of Computing and Informatics University of North Carolina Charlotte (invited).

Wheeler — October 2013. “The evolution and biogeography of influenza.” Universidad Nacional Autónoma de México, Mexico City, Mexico (invited).

Wheeler — October 2013. “The general tree-alignment problem.” Universidad Nacional Autónoma de México, Mexico City, Mexico (invited).

Wheeler — October 2013. “Historical linguistics and a sequence optimization problem: the evolution and biogeography of Uto-Aztecan Languages.” Universidad Nacional Autónoma de México, Mexico City, Mexico (invited).

Janies — September 12, 2013. “Geographic and Genetic Observations of the Spread of Infectious Diseases” International Conference on Genomics in the Americas, Sacramento, CA (invited).

Janies — August 2013. “The Supramap Project: Genomes and Geography for One Health. Use case: H7N9 influenza outbreak in China in 2013.” International Congress of Pathogens at the Human-Animal Interface (ICOPHAI): One Health for Sustainable Development, Porto de Galinhas, Brazil (contributed). Plus two invited lectures on bioinformatics and geography in an international pre-congress course on molecular epidemiology.

Wheeler — August 2013. “Heuristic algorithms, optimality criteria, and computational challenges in the analysis of comparative sequence data.” XXXII Meeting of the Willi Hennig Society, Rostock, Germany.

***Science Documentary Videos***

*When Good Bacteria Go Bad*. http://www.amnh.org/explore/science-bulletins/(watch)/human/documentaries/when-good-bacteria-go-bad/(p)/1

***Teaching Case***

Bacteria Evolving: Tracing the Origins of a MRSA Epidemic. http://www.amnh.org/explore/curriculum-collections/bacteria-evolving-tracing-the-origins-of-a-mrsa-epidemic

American Museum of Natural History

# Bacteria Evolving:

## *Tracing the Origins of a MRSA Epidemic*

TEACHER VERSION

PASSAGE ONE

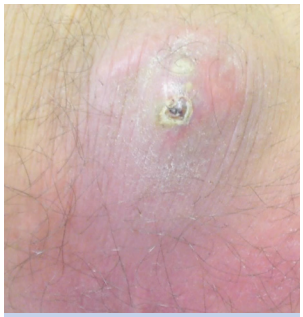
## *What is MRSA?*

We often think of the human body as one organism, but in fact, our bodies are home to a complex ecosystem of microorganisms. Right now more than 8,000 species of bacteria are living in and on your body, and about 1,000 of those are living on your skin. And that's just the number of *species*. All of these bacteria reproduce very quickly, producing a new generation in as little as every twenty minutes. We can only estimate the total number of bacteria in and on your body, but it might be as many as 100 trillion individual organisms.

Along with bacteria, there are other microscopic organisms living on us, including viruses and fungi. Together, these organisms form a ***microbiome***, a microscopic ecosystem. Not only that, within the microbiome are many different microscopic habitats or environments. Each habitat has its own unique population of microorganisms.

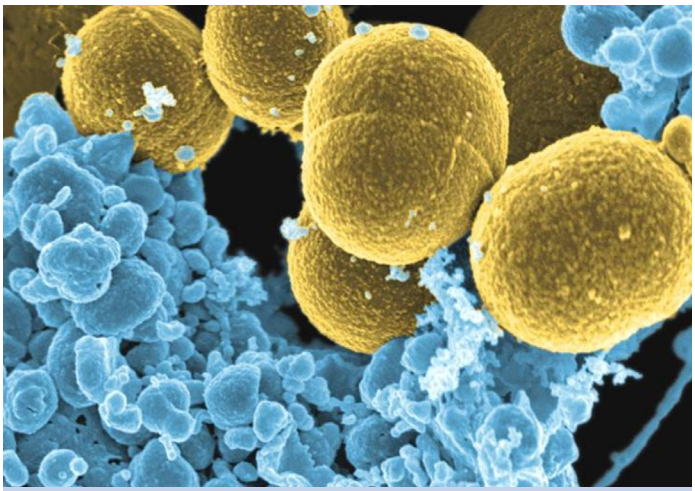
Some bacteria thrive in your hair follicles, others in your sweat glands, others in the relatively moist areas of your face and back.

The relationship between your skin and these microorganisms is quite complex. Your skin is a barrier that keeps these organisms out of your body. At the same time, it also interacts with them. For example, some bacteria on your skin help to "educate" your immune system so it will be better able to recognize dangerous bacteria that invade the body.



A staph infection can occur when the bacteria gets under human skin.

However, sometimes bacteria on our skin can become harmful or even deadly. One example is a group of bacteria called ***MRSA*** or **M**ethicillin-**R**esistant **S**taphylococcus **A**ureus.



*Staphylococcus* gets it name from its round shape. *Coccus* means "little ball" and *staphylo* means "bunch of grapes" (shown here in yellow). That's what staph bacteria look like under the microscope – a bunch of grapes. The word *aureus* refers to its golden pigment.

**What is MRSA?**

*Staphylococcus* or staph is a large group of bacteria. Within that group is a species called ***Staphylococcus aureus*** that has adapted to live on our skin and in our nostrils. About one in three people have staph on their bodies, usually without even knowing it. When bacteria live on us without harming us, it's called *colonization*. However, sometimes bacteria attack healthy cells and make us sick, causing *infection*.

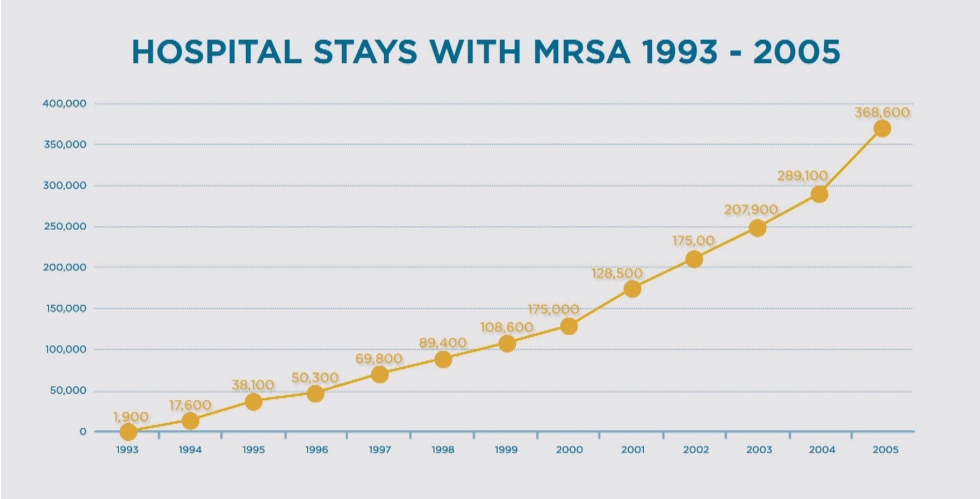
Doctors can usually cure bacterial infections with antibiotic medicines. And that's why MRSA is so dangerous. As you can tell from its name, MRSA is *methicillin-resistant*. Methicillin used to be the antibiotic that was most effective in treating staph infections. But methicillin and other related antibiotics do not kill MRSA. As a result, MRSA infections are very difficult to treat and sometimes fatal.

There are about 100 different types of MRSA, and about one to five percent of people carry MRSA colonies. Within that

CONTINUED

## *What is MRSA?*

group, only a tiny percentage will get a MRSA infection. Like other *S. aureus*, a MRSA bacterium by itself does not make you sick, but can get under your skin and become quite serious.



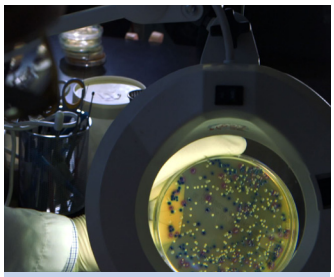

Cases of MRSA reported in U.S. hospitals have skyrocketed over the past two decades, prompting researchers to track the evolution of this rapidly changing epidemic.

### A New Type of MRSA

MRSA bacteria ***strains*** (subtypes within a species) have been around for a long time. In fact, only two years after the introduction of methicillin in the 1950's, the first cases of methicillin-resistant *S. aureus* appeared in hospitals. This is called *hospital-acquired-MRSA*. It seems to spread mainly through hospital equipment and personnel.

Then, in the late 1990's, a new type of MRSA began showing up. The new strains, unlike hospital-acquired-MRSA, were causing infections in healthy people outside of hospitals. This was called *community-acquired-MRSA*. The most common community-acquired-MRSA strain was called USA300. Although it was more susceptible to antibiotics, USA300 was somehow much better at spreading from person to person than hospital-acquired-MRSA.

It's clear why this new type of MRSA was so troubling to doctors and public health workers. They feared its ability to spread outside of hospital settings would lead to dangerous epidemics. And in fact, cases of MRSA, though still relatively rare, have increased dramatically in past decades.



Researchers take a close look at *Staphylococcus aureus* in a petri dish at Columbia University Medical Center in New York City.

Scientists and health workers knew that MRSA had evolved. Somehow, the strain called USA300 had gained the ability to spread more easily from person to person. But how was it able to do that? And how had it gotten this new ability? That was the mystery that scientists set out to solve. If they could find the answer, they just might be able to find a way to stop MRSA from causing serious infections, even fatalities.

### *STOP AND THINK*

***Based on the text:***

- **What are the medical implications of antibiotic resistance?**
- **What factors make USA300, a MRSA strain, of special concern?**
- **What are some of the questions that scientists ask about USA300?**

***Looking ahead:***

- **What kind of data do you think scientists might collect to answer those questions?**

# Bacteria Evolving:

## *Tracing the Origins of a MRSA Epidemic*

STUDENT VERSION

PASSAGE ONE

## *What is MRSA?*

We often think of the human body as one organism, but in fact, our bodies are home to a complex ecosystem of microorganisms. Right now more than 8,000 species of bacteria are living in and on your body, and about 1,000 of those are living on your skin. And that's just the number of *species*. All of these bacteria reproduce very quickly, producing a new generation in as little as every twenty minutes. We can only estimate the total number of bacteria in and on your body, but it might be as many as 100 trillion individual organisms.

Along with bacteria, there are other microscopic organisms living on us, including viruses and fungi. Together, these organisms form a ***microbiome***, a microscopic ecosystem. Not only that, within the microbiome are many different microscopic habitats or environments. Each habitat has its own unique population of microorganisms.

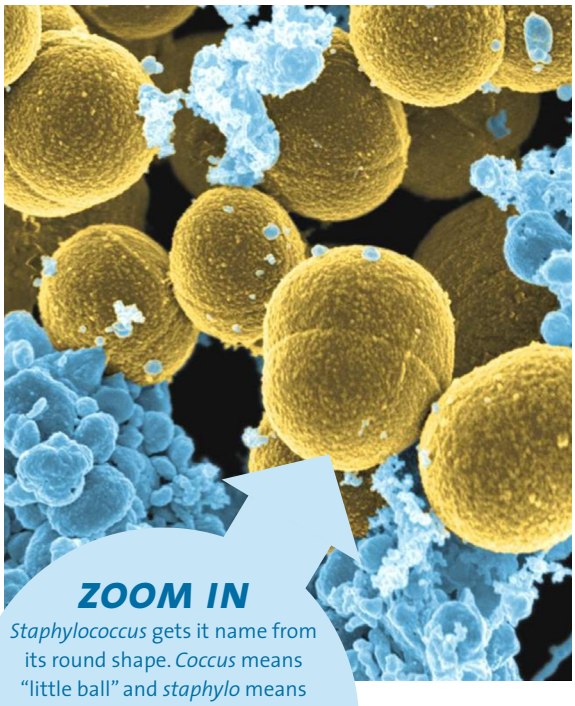
Some bacteria thrive in your hair follicles, others in your sweat glands, others in the relatively moist areas of your face and back.

The relationship between your skin and these microorganisms is quite complex. Your skin is a barrier that keeps these organisms out of your body. At the same time, it also interacts with them. For example, some bacteria on your skin help to "educate" your immune system so it will be better able to recognize dangerous bacteria that invade the body.

However, sometimes bacteria on our skin can become harmful or even deadly. One example is a group of bacteria called ***MRSA*** or **M**ethicillin-**R**esistant **S**taphylococcus **A**ureus.

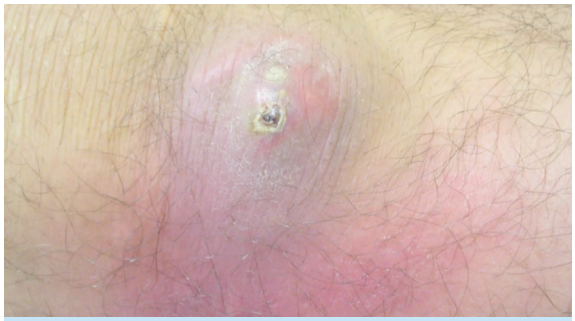
### What is MRSA?

*Staphylococcus* or staph is a large group of bacteria. Within that group is a species called ***Staphylococcus aureus*** that has adapted to live on our skin and in our nostrils. About one in three people have staph on their bodies, usually without even knowing it.



**ZOOM IN**

*Staphylococcus* gets it name from its round shape. *Coccus* means "little ball" and *staphylo* means "bunch of grapes" (shown here in yellow). That's what staph bacteria look like under the microscope – a bunch of grapes. The word *aureus* refers to its golden pigment.



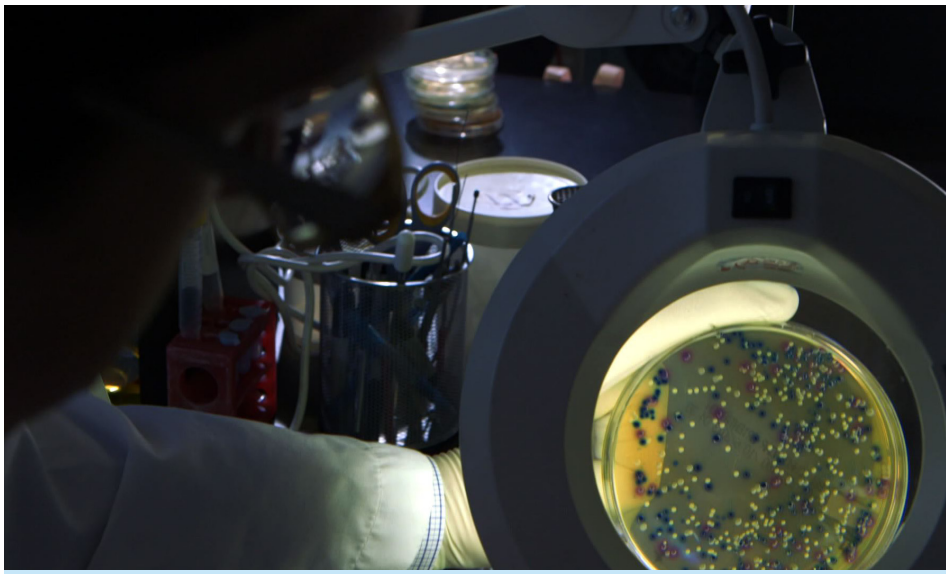
A staph infection can occur when the bacteria gets under human skin.

CONTINUED

### *What is MRSA?*

When bacteria live on us without harming us, it's called *colonization*. However, sometimes bacteria attack healthy cells and make us sick, causing *infection*.

Doctors can usually cure bacterial infections with antibiotic medicines. And that's why MRSA is so dangerous. As you can tell from its name, MRSA is *methicillin-resistant*. Methicillin used to be the antibiotic that was most effective in treating staph infections. But methicillin and other related antibiotics do not kill MRSA. As a result, MRSA infections are very difficult to treat and sometimes fatal.



**UNDER THE MICROSCOPE**
Researchers take a close look at *Staphylococcus aureus* in a petri dish at Columbia University Medical Center in New York City.

There are about 100 different types of MRSA, and about one to five percent of people carry MRSA colonies. Within that group, only a tiny percentage will get a MRSA infection. Like other *S. aureus*, a MRSA bacterium by itself does not make you sick, but can get under your skin and become quite serious.

**A New Type of MRSA**

MRSA bacteria ***strains*** (subtypes within a species) have been around for a long time. In fact, only two years after the introduction of methicillin in the 1950's, the first cases of methicillin-resistant *S. aureus* appeared in hospitals. This is called *hospital-acquired-MRSA*. It seems to spread mainly through hospital equipment and personnel.

Then, in the late 1990's, a new type of MRSA began showing up. The new strains, unlike hospital-acquired-MRSA, were causing infections in healthy people outside of hospitals. This was called

## FAST FACTS

- *Staphylococcus aureus* is the leading cause of surgical wound infections
- MRSA can spread in two ways: between patients and health-care workers in a hospital, or individuals living and working together in close community
- Over 2 million patients in the United States get an infection in the hospital each year

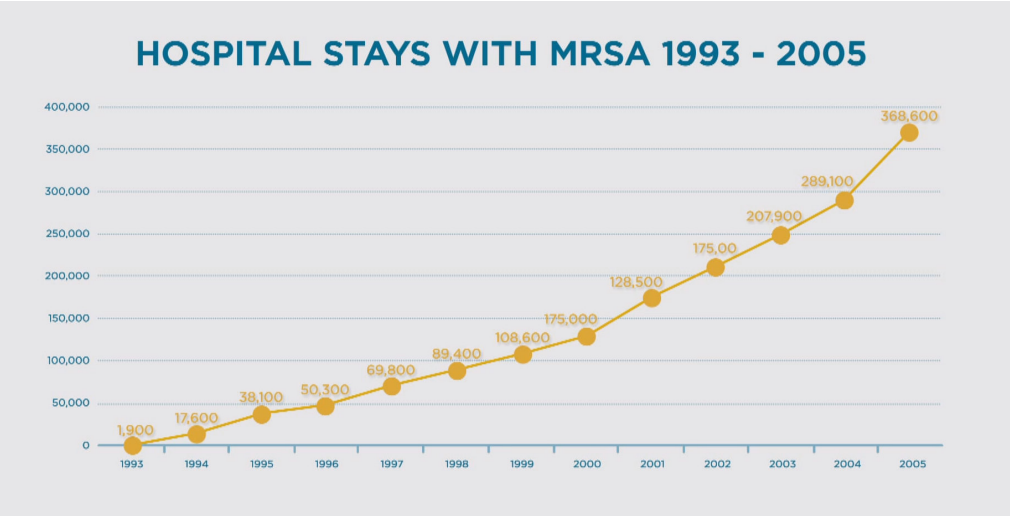
CONTINUED

***What is MRSA?***

*community-acquired-MRSA*. The most common community-acquired-MRSA strain was called USA300. Although it was more susceptible to antibiotics, USA300 was somehow much better at spreading from person to person than hospital- acquired-MRSA.

It's clear why this new type of MRSA was so troubling to doctors and public health workers. They feared its ability to spread outside of hospital settings would lead to dangerous epidemics. And in fact, cases of MRSA, though still relatively rare, have increased dramatically in past decades.

Scientists and health workers knew that MRSA had evolved. Somehow, the strain called USA300 had gained the ability to spread more easily from person to person. But how was it able to do that? And how had it gotten this new ability? That was the mystery that scientists set out to solve. If they could find the answer, they just might be able to find a way to stop MRSA from causing serious infections, even fatalities.




## TRACKING THE EPIDEMIC

Cases of MRSA reported in U.S. hospitals have skyrocketed over the past two decades, prompting researchers to track the evolution of this rapidly changing epidemic.

## STOP AND THINK

***Based on the text:***

- **What are the medical implications of antibiotic resistance?**
- **What factors make USA300, a MRSA strain, of special concern?**
- **What are some of the questions that scientists ask about USA300?**

***Looking ahead:***

- **What kind of data do you think scientists might collect to answer those questions?**

# Bacteria Evolving:

## *Tracing the Origins of a MRSA Epidemic*

TEACHER VERSION

PASSAGE TWO

## *How Did MRSA Evolve?*

What allowed USA300 to thrive outside a hospital setting, and with such ***virulence***? Scientists knew the answer lay in the bacteria's genome, the sum of all its genetic material. Within the bacteria's genome were the clues that could explain how the USA300 strain of MRSA had evolved.

**Different Types of Mutations**

Like all organisms, bacteria can acquire new traits through mutations. Mutations are any change in the sequence of DNA nucleotides within an organism's genome. The main cause of mutations are exposure to foreign chemicals or radiation, errors during DNA replication, and from insertion or deletion of DNA segments. If a mutation is beneficial, it gives the organism an evolutionary advantage by helping it and its descendants to survive in a new environment. This is the process of *natural selection*.

These types of mutations can happen in any organism. In bacteria, DNA can also be acquired through the process of DNA transfer. That means that bacteria can get whole new sets of DNA from other bacteria, creating sudden and dramatic changes in their genome. They can even do this with bacteria that are distantly related, the equivalent of a human being getting new genes from a cat.

This ability to pick up entire sets of foreign DNA is what makes bacteria evolution so sudden and unpredictable. Twenty-five percent or more of a bacterium's genome might be acquired in this way. It gives them a unique way of picking up genes that other bacteria have already tried out. That's a great evolutionary tool.

### Definitions

**DNA:** Deoxyribonucleic acid is the organic molecule that forms the genetic material of an organism. Chromosomes are made of DNA.

**gene:** a section of DNA on a chromosome that encodes, individually or as part of a group of genes, for a specific hereditary trait.

**genome:** the complete genetic material or base sequence of an organism or species.

**genotype:** the genetic makeup of an individual or species.

**phenotype:** the observable characteristics of an organism or species, including its appearance, structure, behaviors.

**selective pressure:** a factor that affects the reproductive success of individuals in a population.

**virulence:** ability of some bacteria to cause serious illness.

CONTINUED

## *How Did MRSA Evolve?*

**Figure 1: BACTERIAL TRANSFORMATION**

**Figure 2: BACTERIAL TRANSDUCTION**

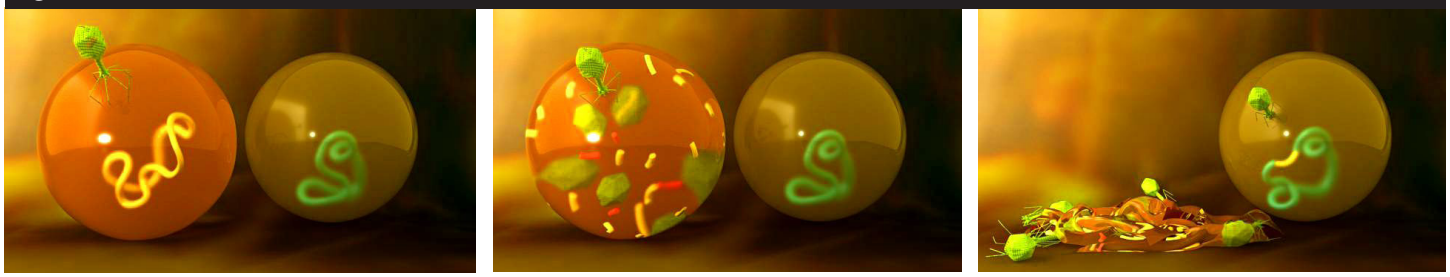
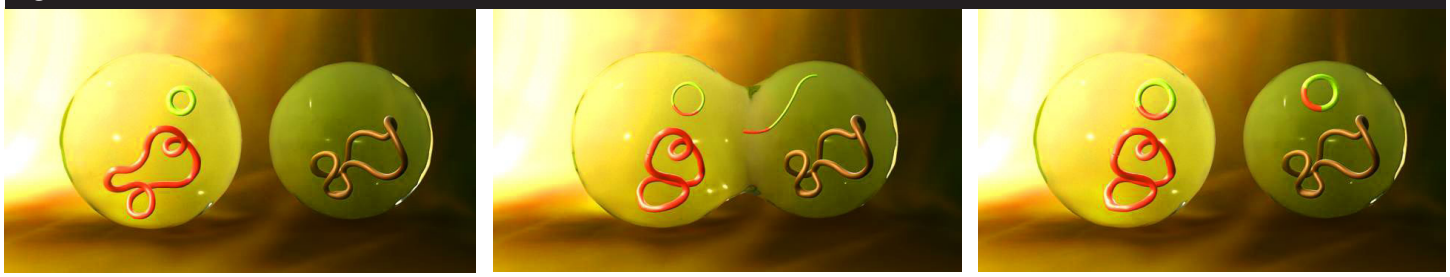


**Figure 3: BACTERIAL CONJUGATION**



DNA transfer in bacteria can happen in several ways.

- Bacteria can recognize loose pieces of DNA that are floating nearby. These pieces might come from a bacterium that ruptured. In a process called ***transformation*** (Figure 1), the bacterium can actively reach out and pull the loose DNA through its membranes.
- Viruses called bacteriophages can enter a bacterium and replicate inside it. At the same time, the virus can pick up DNA from the infected cell, move it over and inject it into another cell. The DNA becomes part of the second organism's genome. This process is called ***transduction*** (Figure 2).
- Bacteria can also trade DNA with each other, in a process called ***conjugation*** (Figure 3). In conjugation, two cells come close to each other and form a bridge or link between them. Then one of the bacteria transfers a bit of DNA to the other.

These still images from an animation show three ways that DNA can be transferred in bacteria.

CONTINUED

## *How Did MRSA Evolve?*

**The Evolution of USA300**

In order to develop ways to combat MRSA infections, scientists needed to study its genetic material. They were especially interested in finding out what made the most virulent of the community-acquired strains of MRSA, USA300, different from the older, hospital-acquired strains.

The genome of USA300 had already been mapped, so the researchers were able to compare the DNA of USA300 with the genomes of other strains of ***S. aureus*** bacteria. When researchers did that, important differences jumped out at them. The USA300 MRSA had genes that had never been seen before in ***S. aureus*** bacteria, including a set of 34 genes called the Arginine Catabolic Mobile Elements (ACME) region. One of those 34 genes is one they called speG.

### How Antibiotics Can Create Drug-Resistant Bacteria

Natural selection explains how the overuse of antibiotics leads to new strains of drug resistant bacteria. In all organisms, members of the same population have slightly different genes or DNA. Because of genetic variation within bacteria populations, some members can be easily killed off by antibiotics, while others will not. If the population is repeatedly exposed to an antibiotic, the ***selective pressure*** in the population will allow the most drug-resistant bacteria to survive. Over time, you have a population in which all the members are drug-resistant.

### *STOP AND THINK*

***Based on the text:***

- **What kind of data are the scientists collecting? How does this compare to your answer from Passage 1?**
- **What are the implications of DNA transfer for the development of antibiotic resistance in bacteria?**
- **How does natural selection explain the development of antibiotic resistance?**

***Looking ahead:***

- **What experiments could the scientists design to find out whether the speG gene is responsible for USA300's unique characteristics?**

# Bacteria Evolving:

## *Tracing the Origins of a MRSA Epidemic*

STUDENT VERSION

PASSAGE TWO

## *How Did MRSA Evolve?*

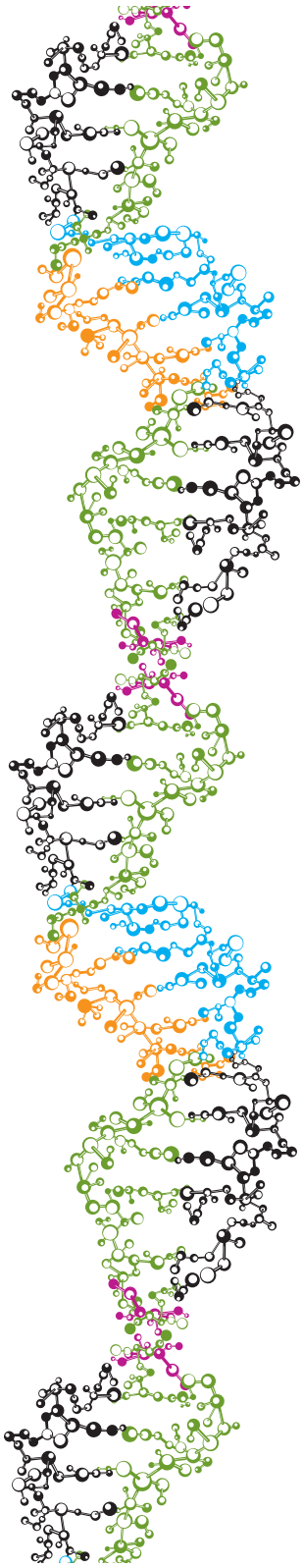
What allowed USA300 to thrive outside a hospital setting, and with such ***virulence***? Scientists knew the answer lay in the bacteria's genome, the sum of all its genetic material. Within the bacteria's genome were the clues that could explain how the USA300 strain of MRSA had evolved.

**Different Types of Mutations**

Like all organisms, bacteria can acquire new traits through mutations. Mutations are any change in the sequence of DNA nucleotides within an organism's genome. The main cause of mutations are exposure to foreign chemicals or radiation, errors during DNA replication, and from insertion or deletion of DNA segments. If a mutation is beneficial, it gives the organism an evolutionary advantage by helping it and its descendants to survive in a new environment. This is the process of *natural selection*.

These types of mutations can happen in any organism. In bacteria, DNA can also be acquired through the process of DNA transfer. That means that bacteria can get whole new sets of DNA from other bacteria, creating sudden and dramatic changes in their genome. They can even do this with bacteria that are distantly related, the equivalent of a human being getting new genes from a cat.

This ability to pick up entire sets of foreign DNA is what makes bacteria evolution so sudden and unpredictable. Twenty-five percent or more of a bacterium's genome might be acquired in this way. It gives them a unique way of picking up genes that other bacteria have already tried out. That's a great evolutionary tool.



### *DEFINITIONS*

**DNA:** Deoxyribonucleic acid is the organic molecule that forms the genetic material of an organism. Chromosomes are made of DNA.

**gene:** a section of DNA on a chromosome that encodes, individually or as part of a group of genes, for a specific hereditary trait.

**genome:** the complete genetic material or base sequence of an organism or species.

**genotype:** the genetic makeup of an individual or species.

**phenotype:** the observable characteristics of an organism or species, including its appearance, structure, behaviors.

**selective pressure:** a factor that affects the reproductive success of individuals in a population.

**virulence:** ability of some bacteria to cause serious illness.

CONTINUED

## *How Did MRSA Evolve?*

**Figure 1: BACTERIAL TRANSFORMATION**

**Figure 2: BACTERIAL TRANSDUCTION**

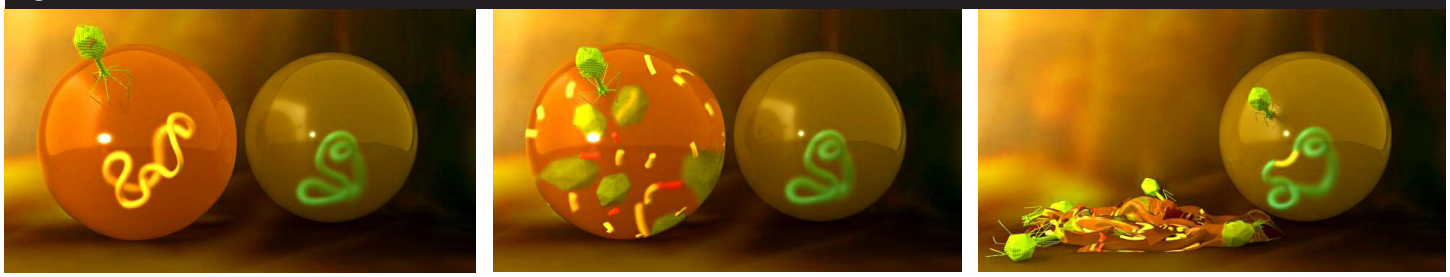
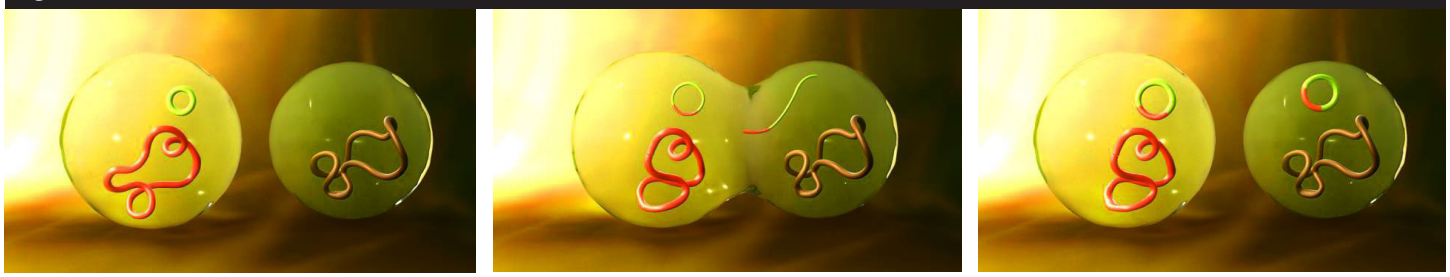


**Figure 3: BACTERIAL CONJUGATION**



DNA transfer in bacteria can happen in several ways.

- Bacteria can recognize loose pieces of DNA that are floating nearby. These pieces might come from a bacterium that ruptured. In a process called ***transformation*** (Figure 1), the bacterium can actively reach out and pull the loose DNA through its membranes.
- Viruses called bacteriophages can enter a bacterium and replicate inside it. At the same time, the virus can pick up DNA from the infected cell, move it over and inject it into another cell. The DNA becomes part of the second organism's genome. This process is called ***transduction*** (Figure 2).
- Bacteria can also trade DNA with each other, in a process called ***conjugation*** (Figure 3). In conjugation, two cells come close to each other and form a bridge or link between them. Then one of the bacteria transfers a bit of DNA to the other.

### *STEP BY STEP*

These still images from an animation show three ways that DNA can be transferred in bacteria.

CONTINUED

## *How Did MRSA Evolve?*

### The Evolution of USA300

In order to develop ways to combat MRSA infections, scientists needed to study its genetic material. They were especially interested in finding out what made the most virulent of the community-acquired strains of MRSA, USA300, different from the older, hospital-acquired strains.

The genome of USA300 had already been mapped, so the researchers were able to compare the DNA of USA300 with the genomes of other strains of ***S. aureus*** bacteria. When researchers did that, important differences jumped out at them. The USA300 MRSA had genes that had never been seen before in ***S. aureus*** bacteria, including a set of 34 genes called the Arginine Catabolic Mobile Elements (ACME) region. One of those 34 genes is one they called speG.

### *STOP AND THINK*

***Based on the text:***

- **What kind of data are the scientists collecting? How does this compare to your answer from Passage 1?**
- **What are the implications of DNA transfer for the development of antibiotic resistance in bacteria?**
- **How does natural selection explain the development of antibiotic resistance?**

***Looking ahead:***

- **What experiments could the scientists design to find out whether the speG gene is responsible for USA300's unique characteristics?**

### *HOW ANTIBIOTICS CREATE DRUG-RESISTANT BACTERIA*

Natural selection explains how the overuse of antibiotics leads to new strains of drug resistant bacteria. In all organisms, members of the same population have slightly different genes or DNA. Because of genetic variation within bacteria populations, some members can be easily killed off by antibiotics, while others will not. If the population is repeatedly exposed to an antibiotic, the ***selective pressure*** in the population will allow the most drug-resistant bacteria to survive. Over time, you have a population in which all the members are drug-resistant.

# Bacteria Evolving:

*Tracing the Origins of a MRSA Epidemic*

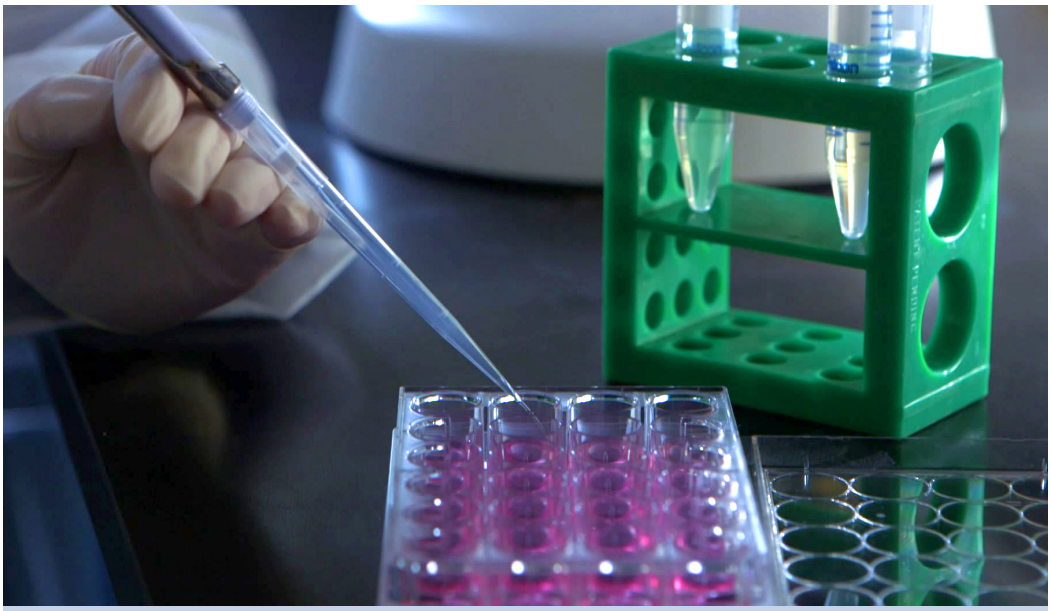
TEACHER VERSION

PASSAGE THREE

## *Testing the Hypothesis*

Scientists' initial hypothesis was that the speG gene was the cause of spermidine-resistence in USA300. Spermidine is a natural antibiotic given off by your skin when you have a cut or abrasion. They reasoned that the speG gene in USA300 makes it completely resistant to the killing power of spermidine.

Now they had to test this hypothesis. They wondered: What if they could disable the speG gene in USA300? What would happen then? Would USA300 still be resistant to spermidine?



At the start of their experiment, scientists pipette knock-out samples of the speG gene onto live human skin cells growing in vials at Columbia University Medical Center in New York City.

To conduct this test, the scientists studied the effect of spermidine both on a "wild type" USA300 and on a "knockout" USA300, in which the original speG gene is either replaced by a non-functioning mutant copy of the gene, or the gene is deleted from the genome altogether. They expected that the modified version of USA300 would not be able to alter or neutralize spermidine.

First they had to prepare the knockout USA300. How could they get the bacteria to take in the foreign DNA with the modified gene? Well, as the scientists know, bacteria are very good at adopting foreign DNA into their genome. They do it all the time. They researchers decided to use the bacteria's own process of transformation. They placed USA300 bacteria in a medium that contained the mutant speG gene. Then they used small electric shocks to stimulate the USA300 bacteria to open pores in their outside cell membranes and take in the foreign DNA. This way, they successfully got some of the bacteria to replace their original speG gene with a non-functioning version.

This allowed them to run side-by-side tests. They recreated the conditions of a skin infection by growing human skin cells in a culture. To each sample they added some spermidine and then introduced the two versions of USA300, the wild type and the knockout. The results supported their hypothesis. The knockout bacteria fared much worse than the wild group with functioning speG genes. Without a defense against the spermidine, the proportion of mutant USA300 bacteria that survived was much smaller. The difference was clear: it was the speG gene that gives USA300 protection from killing by human skin. They also found that the speG gene allows USA300 to better attach itself to human skin.

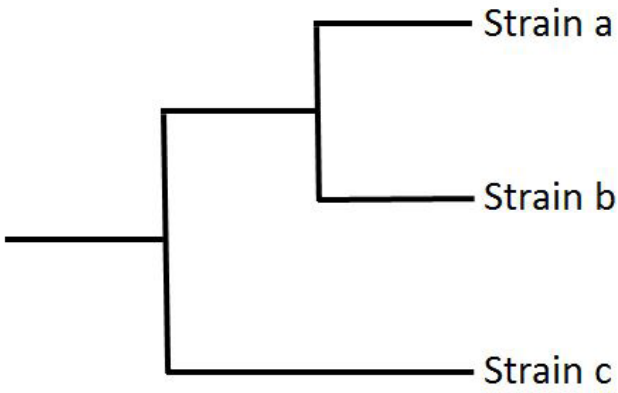
Next, scientists wondered when USA300 first acquired the speG gene. To find out they created a phylogenetic tree of MRSA strains. A phylogenetic tree, or cladogram, is a branching diagram that shows the relationships among organisms based on comparisons of physical or molecular characteristics. The diagram indicates which groups are most closely related. The order of branching points on the tree can be used

CONTINUED

### *Testing the Hypothesis*

to infer the order of historical events in a particular lineage. Scientists used this technique that the acquisition of the speG in the USA300 lineage probably occurred around 1997, which coincides with the first outbreaks of community-acquired MRSA.

But there was one more piece to the puzzle. How did USA300 get the speG gene? Where did it come from? They knew the answer probably lay somewhere in the vast microbiome that exists on human skin.




This simple phylogenetic tree shows that strain a and strain b are more closely related to each other than either is to strain c.

### *STOP AND THINK*

***Based on the text:***

- **What experiments did the scientists design to find out whether the speG gene is responsible for USA300's unique characteristics? Create diagram that shows the design of the experiment. How does this compare to your answer from Passage 2?**
- **What question did the scientists answer by building a phylogenetic tree?**

***Looking ahead:***

- **If USA300 acquired the speG gene through DNA transfer, what steps might the scientists take in order to find the source of that gene?scientists design to find out whether the speG gene is responsible for USA300's unique characteristics?**

# Bacteria Evolving:

## *Tracing the Origins of a MRSA Epidemic*

STUDENT VERSION

PASSAGE THREE

## *Testing the Hypothesis*

Scientists' initial hypothesis was that the speG gene was the cause of spermidine-resistence in USA300. Spermidine is a natural antibiotic given off by your skin when you have a cut or abrasion. They reasoned that the speG gene in USA300 makes it completely resistant to the killing power of spermidine.

Now they had to test this hypothesis. They wondered: What if they could disable the speG gene in USA300? What would happen then? Would USA300 still be resistant to spermidine?

***IN THE LABORATORY***

At the start of their experiment, scientists pipette knock-out samples of the speG gene onto live human skin cells growing in vials at Columbia University Medical Center in New York City.

To conduct this test, the scientists studied the effect of spermidine both on a "wild type" USA300 and on a "knockout" USA300, in which the original speG gene is either replaced by a non-functioning mutant copy of the gene, or the gene is deleted from the genome altogether. They expected that the modified version of USA300 would not be able to alter or neutralize spermidine.

First they had to prepare the knockout USA300. How could they get the bacteria to take in the foreign DNA with the modified gene? Well, as the scientists know, bacteria are very good at adopting foreign DNA into their genome. They do it all the time. They researchers decided to use the bacteria's own process of transformation. They placed USA300 bacteria in a medium that contained the mutant speG gene. Then they used small electric shocks to stimulate the USA300 bacteria to open pores in their outside cell membranes and take in the foreign DNA. This way, they successfully got some of the bacteria to replace their original speG gene with a non-functioning version.

This allowed them to run side-by-side tests. They recreated the conditions of a skin infection by growing human skin cells in a culture. To each sample they added some spermidine and then introduced the two versions of USA300, the wild type and the knockout. The results supported their hypothesis. The knockout

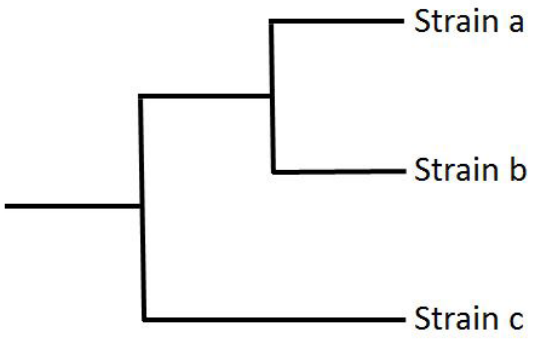
CONTINUED

***Testing the Hypothesis***

bacteria fared much worse than the wild group with functioning speG genes. Without a defense against the spermidine, the proportion of mutant USA300 bacteria that survived was much smaller. The difference was clear: it was the speG gene that gives USA300 protection from killing by human skin. They also found that the speG gene allows USA300 to better attach itself to human skin.

Next, scientists wondered when USA300 first acquired the speG gene. To find out they created a phylogenetic tree of MRSA strains. A phylogenetic tree, or cladogram, is a branching diagram that shows the relationships among organisms based on comparisons of physical or molecular characteristics. The diagram indicates which groups are most closely related. The order of branching points on the tree can be used to infer the order of historical events in a particular lineage. Scientists used this technique that the acquisition of the speG in the USA300 lineage probably occurred around 1997, which coincides with the first outbreaks of community-acquired MRSA.

But there was one more piece to the puzzle. How did USA300 get the speG gene? Where did it come from? They knew the answer probably lay somewhere in the vast microbiome that exists on human skin.




### CLOSE RELATIVES

This simple phylogenetic tree shows that strain a and strain b are more closely related to each other than either is to strain c.

### STOP AND THINK

***Based on the text:***

- **What experiments did the scientists design to find out whether the speG gene is responsible for USA300's unique characteristics? Create diagram that shows the design of the experiment. How does this compare to your answer from Passage 2?**
- **What question did the scientists answer by building a phylogenetic tree?**

***Looking ahead:***

- **If USA300 acquired the speG gene through DNA transfer, what steps might the scientists take in order to find the source of that gene?scientists design to find out whether the speG gene is responsible for USA300's unique characteristics?**

# Bacteria Evolving:

## *Tracing the Origins of a MRSA Epidemic*

TEACHER VERSION

PASSAGE FOUR

## *The Human Microbiome*

To solve the mystery of the origins of USA300, researchers looked to the microbiome that lives on human skin. They knew that within this vast, microscopic ecosystem lives a network of organisms that compete with each other, co-exist with each other and are constantly evolving. If USA300 acquired the speG gene, it was most likely from some other bacterium or organism that is part of this ecosystem. And that is exactly what they found. USA300 had gotten its new abilities from another staph species, *Staphylococcus epidermidis*.

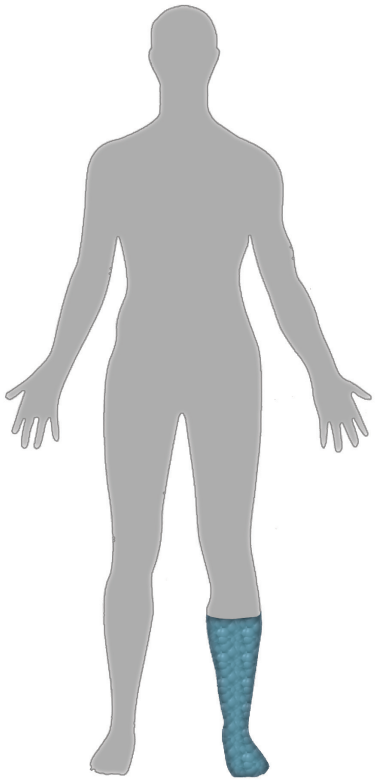
**The Source: *S. epidermidis***

*Staph epidermidis* is a type of *Staphylococcus* that is closely related to *Staphylococcus aureus*. Like *S. aureus, S. epidermidis* has adapted to live on human skin. Although it's quite common, it usually causes no health problems. One reason that *S. epidermidis* can colonize human skin so effectively is that it has the ACME region, the mobile element of DNA made of 34 genes. This is the set of genes that includes the speG gene.

Thanks to its genome, *S. epidermidis* is resistant to methicillin and related antibiotics. However, *S. epidermidis* rarely infects us, so the fact that it is drug resistant is usually not a problem. But what seems to have happened is that *S. epidermidis* bacteria were living in close contact with *S. aureus* bacteria on human skin and at some point *S. epidermidis* transfered the ACME DNA, including the speG gene, to *S. aureus*. This might have made it possible for *S. aureus* to spread more easily from person to person, and made MRSA infections harder to fight.

**Co-Evolving with Humans**

*S. epidermidis* and *S. aureus* are just two of the many microorganisms that colonize human skin. In addition to thousands of types of bacteria, our skin is home to viruses, fungi, mites and other microscopic organisms. These microorganisms are adapted to live in a variety of different environments in and on our bodies. In addition, the microbiome varies from person to person. In other words, different people are home to different sets of microorganisms.



Only 10% of you is actually you! There are ten times more bacterial cells than human cells in your body. If all cells were the same size, your human cells would fit in your foot and lower leg, as illustrated here. But in fact, bacteria cells are much smaller than human cells.

Most of these organisms are not invaders. They have evolved over millions of years to live with human beings. For example, most bacteria cannot survive direct sunlight, but many of the bacteria that live on us have pigments that protect them from ultraviolet light. At the same time, humans have evolved to live with bacteria. We have *co-evolved* or evolved together

CONTINUED

### *The Human Microbiome*

and formed complex relationships.

Scientists are still trying to understand what roles these organisms play, if any, but it seems that most are benign or even beneficial to us. Scientists call these *commensal* bacteria (if they are harmless) or *mutualistic* (if they offer a benefit). For example, some types of *Staphylococcus* produce fatty acids that inhibit the growth of fungi and yeast on our skin.

Sometimes bacteria that are harmless or even beneficial in one place can become harmful in another. *Propionibacterium acnes* lives on the skin but if it becomes trapped in a hair follicle, it causes inflammation and acne. *S. epidermidis* is usually harmless but it can travel into the body on catheters and other medical equipment and cause infection.

**Competition in the Biome**

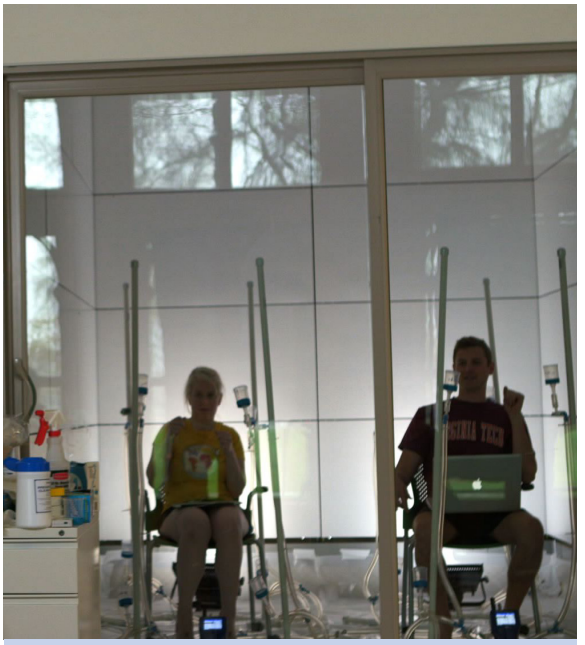
At the same time, all of these microorganisms within the microbiome are competing with each other, just as organisms do in any ecosystem, and bacteria are just one part of these interactions. For example, fungi compete with bacteria for space and resources. Tiny mites that live in our pores eat the fungi. Some of this competition is beneficial to us because our resident microorganisms fight off potential invaders.

The competition between fungi and bacteria is why we have antibiotic medicines. These anti-bacterial chemicals were not invented by human beings but were developed by fungi as they competed with bacteria over millions of years of evolution. Scientists discovered these chemicals and learned how to produce them. But now we may be overusing these chemical weapons and destroying helpful bacteria along with the harmful.

**Being Human Means Having a Microbiome**

Although hand sanitizers and antibiotics have saved millions of lives, their overuse can wreak havoc on our microbiome.

We are covered in bacteria and other microorganisms from the time we are born. And there is mounting evidence that a healthy microbiome is essential for a strong immune system. Some sci-



*Human bacterial cloud*

How do scientists study the human microbiome? How can they get a good picture of which organisms are residents of our bodies? One way is to create a sterile environment or "clean room," place a person in that room, and then find out which microorganisms have been brought in by the human being.

James Meadows, a researcher at the University of Oregon, helped construct a "clean room" experiment. Scientists sealed off a room, sterilized it as much as possible, filtered the air coming in, and then divided it into two chambers. One chamber was kept empty while a person entered the other chamber and sat down.

The scientists found that after repeating the experiment multiple times, they could always tell which chamber the person had entered by measuring the bacteria that had come off his or her body. Not only could they always tell which chamber had been occupied, they could also identify different people from their bacterial "clouds."

CONTINUED

***The Human Microbiome***

entists think that preventing infants from developing a balanced microbiome through lack of exposure to microorganisms might lead to increased allergies, asthma, eczema and other health problems.

Of course, we have to find ways to stop dangerous bacteria like MRSA and other infectious organisms. But as scientists like James Meadows point out, we have to learn to do this in a selective, balanced way. A blanket approach to fighting bacteria creates health risks we still don't understand.

So rather than think of all bacteria as dangerous or harmful, we have to understand that they have always been with us and are an essential part of our microbiome, which is an essential part of our bodies.

## *STOP AND THINK*

***Based on the text:***

- **Describe how the scientists at the University of Oregon study the human microbiome?**
- **What are the benefits of having a diversity of microorganisms living in and on our bodies?**

***Wrap-up:***

- **Consider all four passages and the science practices listed here. By giving specific examples, explain how the scientists in the videos and reading passages apply each of the practices.**

## *PRACTICES FOR K-12 SCIENCE CLASSROOMS from NGSS*

- **Asking questions**
- **Developing and using models**
- **Planning and carrying out investigations**
- **Analyzing and interpreting data**
- **Using mathematics and computational thinking**
- **Constructing explanations**
- **Engaging in argument from evidence**
- **Obtaining, evaluating, and communicating information**

# Bacteria Evolving:

## *Tracing the Origins of a MRSA Epidemic*

STUDENT VERSION

PASSAGE FOUR

## *The Human Microbiome*

To solve the mystery of the origins of USA300, researchers looked to the microbiome that lives on human skin. They knew that within this vast, microscopic ecosystem lives a network of organisms that compete with each other, co-exist with each other and are constantly evolving. If USA300 acquired the speG gene, it was most likely from some other bacterium or organism that is part of this ecosystem. And that is exactly what they found. USA300 had gotten its new abilities from another staph species, *Staphylococcus epidermidis*.

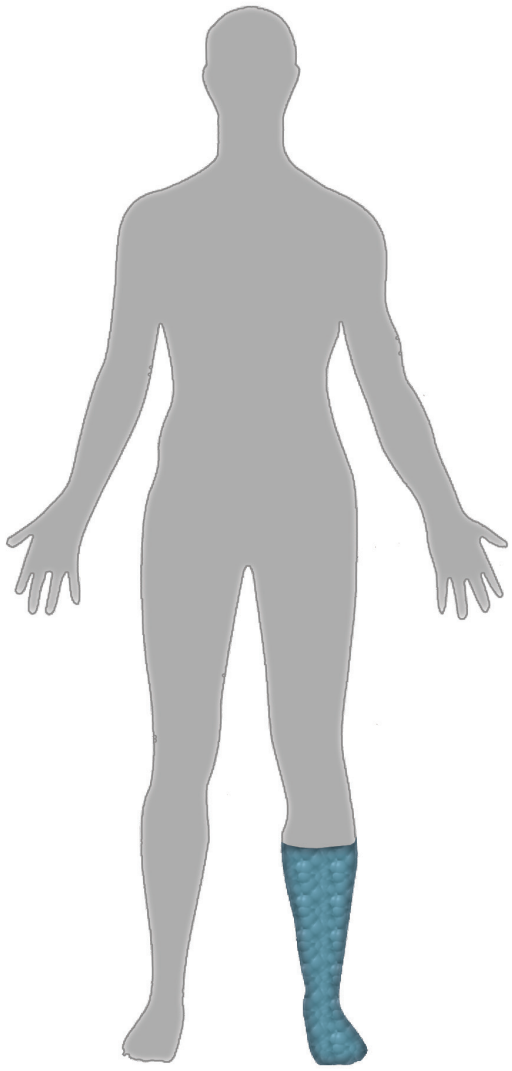
**The Source: *S. epidermidis***

*Staph epidermidis* is a type of *Staphylococcus* that is closely related to *Staphylococcus aureus*. Like *S. aureus, S. epidermidis* has adapted to live on human skin. Although it's quite common, it usually causes no health problems. One reason that *S. epidermidis* can colonize human skin so effectively is that it has the ACME region, the mobile element of DNA made of 34 genes. This is the set of genes that includes the speG gene.

Thanks to its genome, *S. epidermidis* is resistant to methicillin and related antibiotics. However, *S. epidermidis* rarely infects us, so the fact that it is drug resistant is usually not a problem. But what seems to have happened is that *S. epidermidis* bacteria were living in close contact with *S. aureus* bacteria on human skin and at some point *S. epidermidis* transfered the ACME DNA, including the speG gene, to *S. aureus*. This might have made it possible for *S. aureus* to spread more easily from person to person, and made MRSA infections harder to fight.

**Co-Evolving with Humans**

*S. epidermidis* and *S. aureus* are just two of the many microorganisms that colonize human skin. In addition to thousands of types of bacteria, our skin is home to viruses, fungi, mites and other microscopic organisms. These microorganisms are adapted to live in a variety of different environments in and on



**ONLY 10% OF YOU IS ACTUALLY YOU!**

There are ten times more bacterial cells than human cells in your body. Bacteria cells are much smaller than human cells, but if all cells were the same size, your human cells would fit in your foot and lower leg, as illustrated here.

CONTINUED

### *The Human Microbiome*

our bodies. In addition, the microbiome varies from person to person. In other words, different people are home to different sets of microorganisms.

Most of these organisms are not invaders. They have evolved over millions of years to live with human beings. For example, most bacteria cannot survive direct sunlight, but many of the bacteria that live on us have pigments that protect them from ultraviolet light. At the same time, humans have evolved to live with bacteria. We have *co-evolved* or evolved together and formed complex relationships.

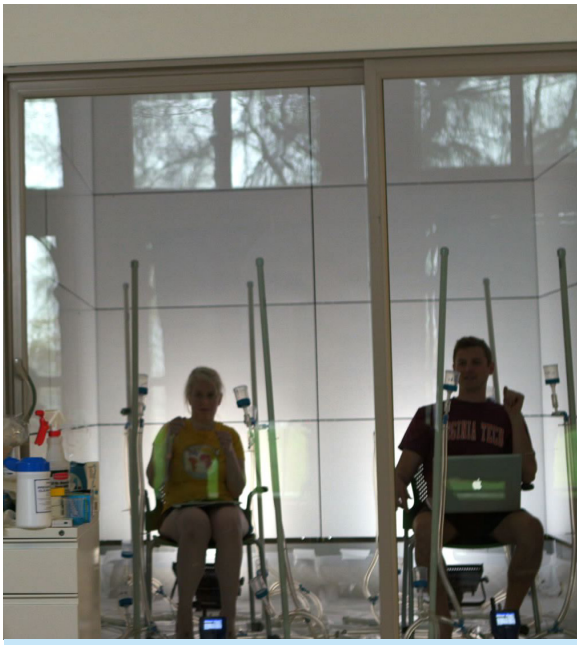
Scientists are still trying to understand what roles these organisms play, if any, but it seems that most are benign or even beneficial to us. Scientists call these *commensal* bacteria (if they are harmless) or *mutualistic* (if they offer a benefit). For example, some types of *Staphylococcus* produce fatty acids that inhibit the growth of fungi and yeast on our skin.

Sometimes bacteria that are harmless or even beneficial in one place can become harmful in another. *Propionibacterium acnes* lives on the skin but if it becomes trapped in a hair follicle, it causes inflammation and acne. *S. epidermidis* is usually harmless but it can travel into the body on catheters and other medical equipment and cause infection.

**Competition in the Biome**

At the same time, all of these microorganisms within the microbiome are competing with each other, just as organisms do in any ecosystem, and bacteria are just one part of these interactions. For example, fungi compete with bacteria for space and resources. Tiny mites that live in our pores eat the fungi. Some of this competition is beneficial to us because our resident microorganisms fight off potential invaders.

The competition between fungi and bacteria is why we have antibiotic medicines. These anti-bacterial chemicals were not invented by human beings but were developed by fungi as they competed with bacteria over millions of years of evolution. Scientists discovered these chemicals and learned how to produce



## HUMAN BACTERIAL CLOUD

How do scientists study the human microbiome? How can they get a good picture of which organisms are residents of our bodies? One way is to create a sterile environment or "clean room," place a person in that room, and then find out which microorganisms have been brought in by the human being.

James Meadows, a researcher at the University of Oregon, helped construct a "clean room" experiment. Scientists sealed off a room, sterilized it as much as possible, filtered the air coming in, and then divided it into two chambers. One chamber was kept empty while a person entered the other chamber and sat down.

The scientists found that after repeating the experiment multiple times, they could always tell which chamber the person had entered by measuring the bacteria that had come off his or her body. Not only could they always tell which chamber had been occupied, they could also identify different people from their bacterial "clouds."

CONTINUED

***The Human Microbiome***

them. But now we may be overusing these chemical weapons and destroying helpful bacteria along with the harmful.

**Being Human Means Having a Microbiome**

Although hand sanitizers and antibiotics have saved millions of lives, their overuse can wreak havoc on our microbiome.

We are covered in bacteria and other microorganisms from the time we are born. And there is mounting evidence that a healthy microbiome is essential for a strong immune system. Some scientists think that preventing infants from developing a balanced microbiome through lack of exposure to microorganisms might lead to increased allergies, asthma, eczema and other health problems.

Of course, we have to find ways to stop dangerous bacteria like MRSA and other infectious organisms. But as scientists like James Meadows point out, we have to learn to do this in a selective, balanced way. A blanket approach to fighting bacteria creates health risks we still don't understand.

So rather than think of all bacteria as dangerous or harmful, we have to understand that they have always been with us and are an essential part of our microbiome, which is an essential part of our bodies.

## STOP AND THINK

***Based on the text:***

- **Describe how the scientists at the University of Oregon study the human microbiome?**
- **What are the benefits of having a diversity of microorganisms living in and on our bodies?**

***Wrap-up:***

- **Consider all four passages and the science practices listed below. By giving specific examples, explain how the scientists in the videos and reading passages apply each of the practices.**
  - **Asking questions**
  - **Developing and using models**
  - **Planning and carrying out investigations**
  - **Analyzing and interpreting data**
  - **Using mathematics and computational thinking**
  - **Constructing explanations**
  - **Engaging in argument from evidence**
  - **Obtaining, evaluating, and communicating information**