



## 16S rRNA gene-based Microbiome Pipeline

# Chimera detection

By Jon Jongsik Chun  
Professor, Seoul National University  
CEO, ChunLab, Inc.



Version FEB-04-2019

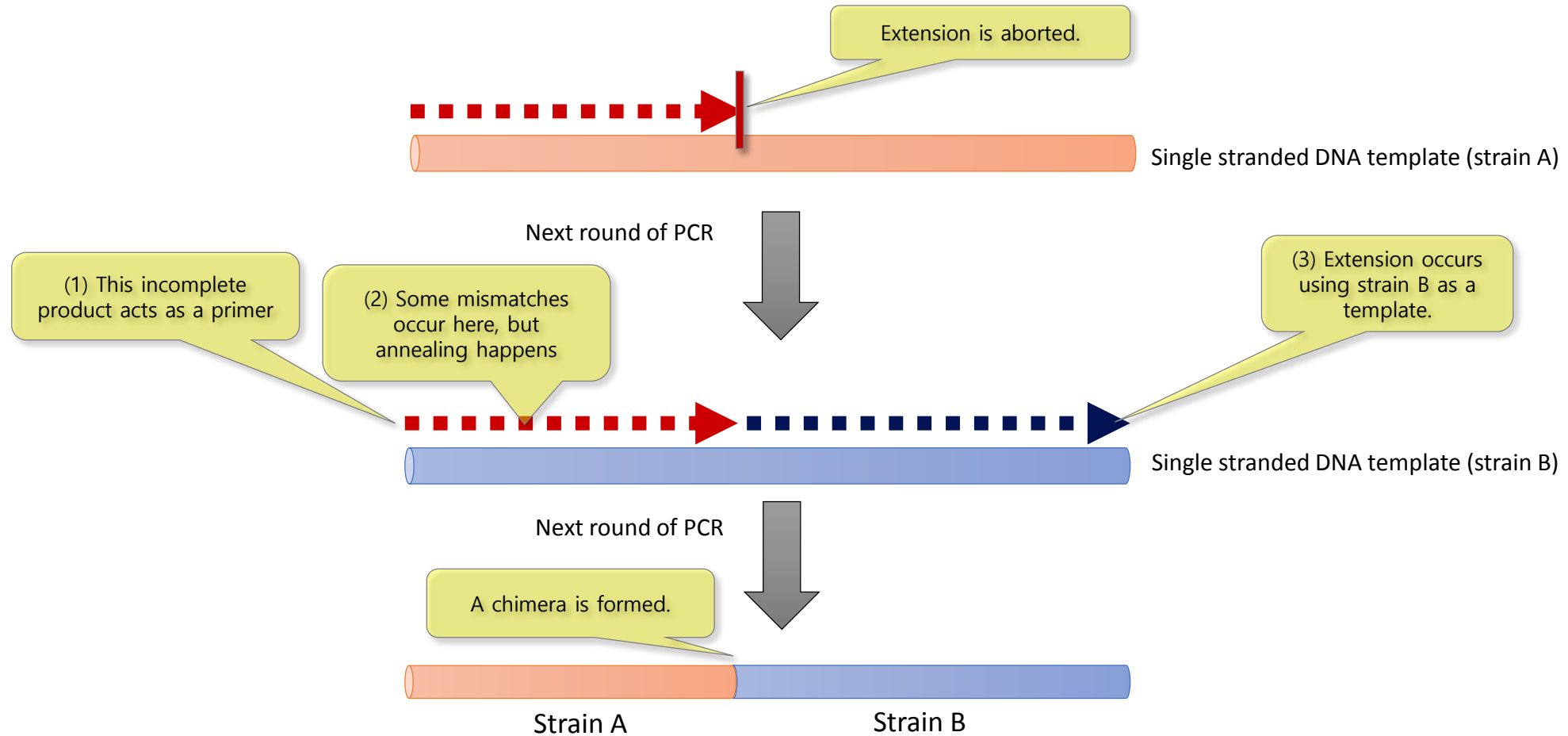
[www.chunlab.com](http://www.chunlab.com)

All lectures are available at  
[https://help.ezbiocloud.net/chun\\_lecture\\_kor/](https://help.ezbiocloud.net/chun_lecture_kor/)

© ChunLab, Inc.



# How do chimeric amplicons form during PCR?



# Typical example of a chimeric amplicon

454 sequence from human skin (V1V3 region)

```
GATGAACGCTGGCGGCGTGCCTAATACATGCAAGTCGAGCGAACAGACGAGGA  
GCTTGCTCCTCTGACGTTAGCGGCGGACGGGTGAGTAACACGTGGATAACCTACC  
TATAAGACTGGGATAACTTCGGGAAACCGGAGCTAATACCGGATAATATATTGAA  
CTGCATGGTTCAATAGTGAAAGACGGTTTTGCTGTCACCTATAGATGGATCCGCG  
CCGCATTAGCTAGTTGGTAAGGTAACGGCTTACCAAGGCAACGATGCGTAGCCG  
ACCTGAGAGGGTGATCGGCCACACTGGAAGTGAACACGGTCCAGACTCCTACG  
GGAGGCAGCAGTGGGGAATATTGCACAATGGGCGGAAGCCTGATGCAGCAACG  
CCGCGTGCGGGATGACGGCCTTCGGGTTGTAAACCGCTTTCGCCTGTGACGAAG  
CGTGAGTGACGGTAATGGGTAAGAAGCACCGGCTAACTACGTG
```



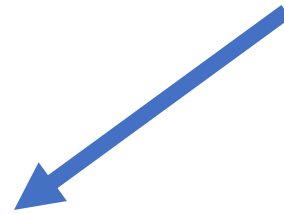
Identify the sequence at <https://www.ezbiocloud.net/>



# Typical example of a chimeric amplicon

454 sequence from human skin (V1V3 region)

```
GATGAACGCTGGCGGCGTGCCTAATACATGCAAGTCGAGCGAACAGACGAGGA  
GCTTGCTCCTCTGACGTTAGCGGCGGACGGGTGAGTAACACGTGGATAACCTACC  
TATAAGACTGGGATAACTTCGGGAAACCGGAGCTAATACCGGATAATATATTGAA  
CTGCATGGTTCAATAGTGAAAGACGGTTTTGCTGTCACCTATAGATGGATCCGCG  
CCGCATTAGCTAGTTGGTAAGGTAACGGCTTACCAAGGCAACGATGCGTAGCCG  
ACCTGAGAGGGTGATCGGCCACACTGGAAGTGAACACGTCAGACTCCTACG  
GGAGGCAGCAGTGGGAATATTGCACAATGGGCGGAAGCCTGATGCAGCAACG  
CCGCGTGCGGGATGACGGCCTTCGGGTTGTAAACCGCTTTCGCCTGTGACGAAG  
CGTGAGTGACGGTAATGGGTAAGAAGCACCGGCTAACTACGTG
```



99.7% to *Staphylococcus epidermidis*

(Firmicutes  
;Bacilli;Bacillales;  
Staphylococcaceae;Staphylococcus)

100% to *Propionibacterium acnes*

(Actinobacteria  
;Actinobacteria;  
Propionibacteriales;Propionibacteriaceae;  
Propionibacterium)

# Chimera formation during PCR amplification

- **Chimera formation is inevitable during 16S rRNA gene PCR of metagenomic samples.**
- **Unfiltered chimeric amplicons will be recognized as new species or OTU, which affect the subsequent statistical analyses on alpha and beta diversity.**



*Sequence analysis*

Advance Access publication June 23, 2011

## **UCHIME improves sensitivity and speed of chimera detection**

Robert C. Edgar<sup>1,\*</sup>, Brian J. Haas<sup>2</sup>, Jose C. Clemente<sup>3</sup>, Christopher Quince<sup>4</sup>  
and Rob Knight<sup>3</sup>

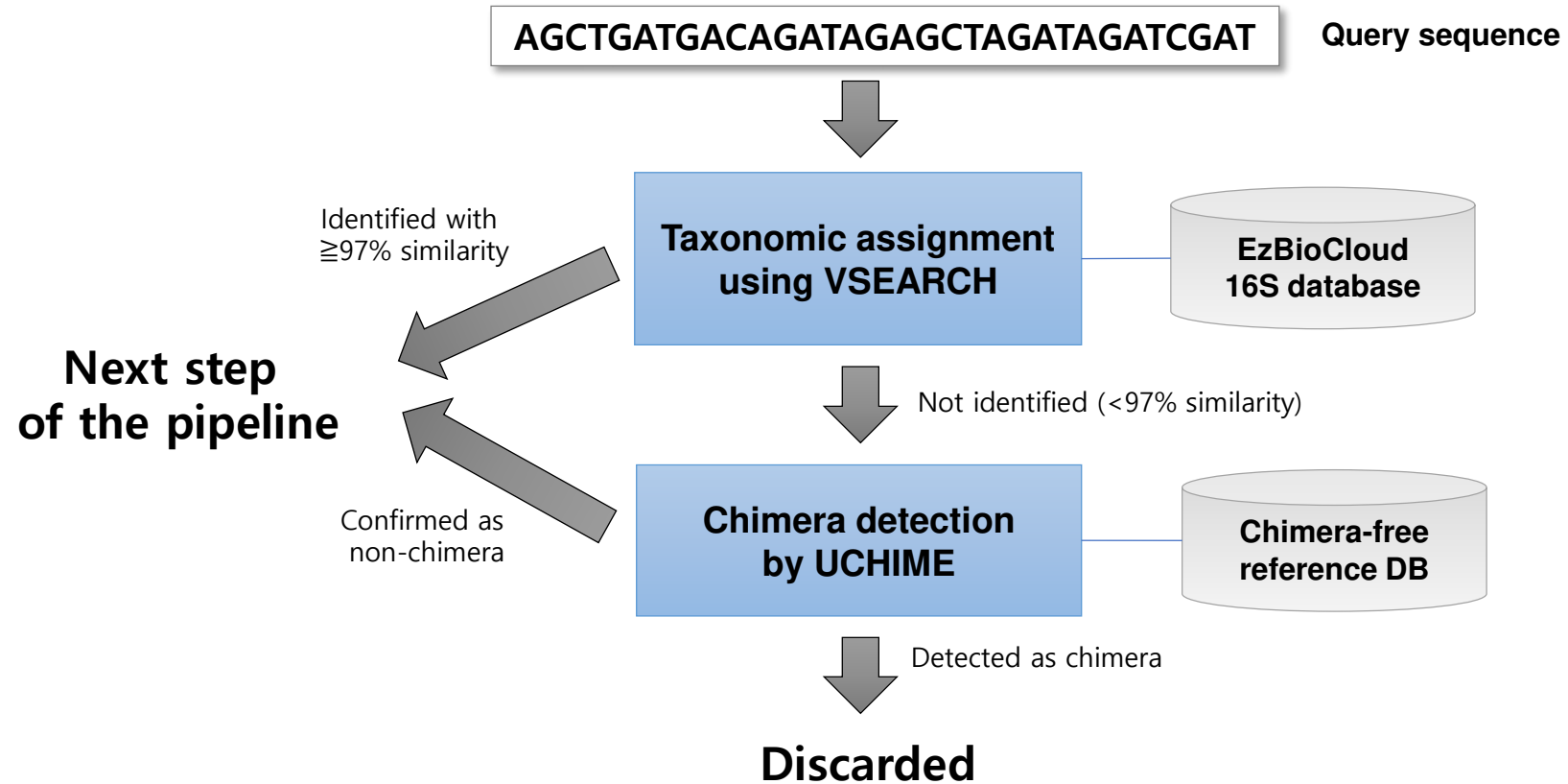
<sup>1</sup>Tiburon, CA, USA, <sup>2</sup>Genome Sequencing and Analysis Program, The Broad Institute, Cambridge, MA 02142,

<sup>3</sup>Department of Chemistry and Biochemistry, University of Colorado, Boulder, CO 80309, USA and <sup>4</sup>School of Engineering, University of Glasgow, Glasgow G12 8LT, UK

Associate Editor: Martin Bishop



# Filtering chimeric sequences in the EzBioCloud pipeline

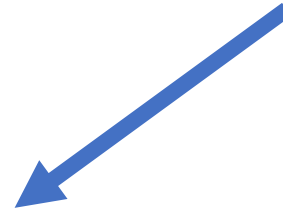




# Example of a chimeric sequence

454 sequence from human skin (V1V3 region)

```
GATGAACGCTGGCGGCGTGCCTAATACATGCAAGTCGAGCGAACAGACGAGGA  
GCTTGCTCCTCTGACGTTAGCGGCGGACGGGTGAGTAACAGTGGATAACCTACC  
TATAAGACTGGGATAACTTCGGGAAACCGGAGCTAATACCGGATAATATATTGAA  
CTGCATGGTTCAATAGTGAAAGACGGTTTTGCTGCACTTATAGATGGATCCGCG  
CCGCATTAGCTAGTTGGTAAGGTAACGGCTTACCAAGGCAACGATGCGTAGCCG  
ACCTGAGAGGGTGATCGGCCACACTGGAAGTGAACACGGTCCAGACTCCTACG  
GGAGGCAGCAGTGGGGAATATTGCACAATGGGCGGAAGCCTGATGCAGCAACG  
CCGCGTGCGGGATGACGGCCTTCGGGTTGTAAACCGCTTTCGCCTGTGACGAAG  
CGTGAGTGACGGTAATGGGTAAGAAGCACCGGCTAACTACGTG
```

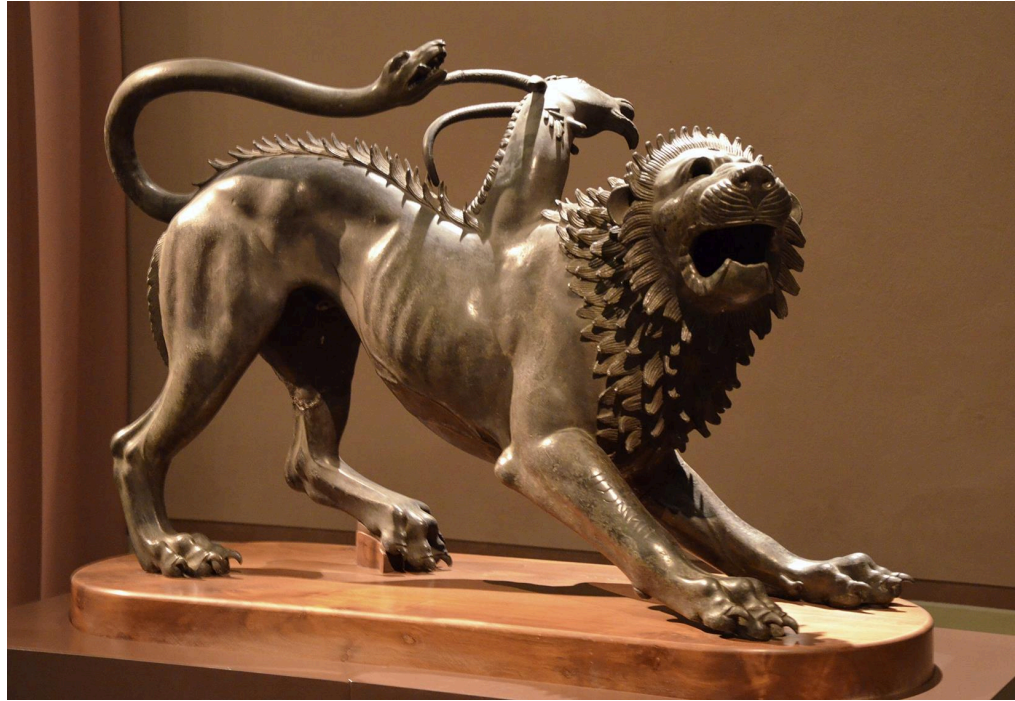


**99.7% to *Staphylococcus epidermidis***  
**(Firmicutes; Bacilli; Bacillales;  
Staphylococcaceae; Staphylococcus)**

**100% to *Propionibacterium acnes***  
**(Actinobacteria; Actinobacteria;  
Propionibacteriales; Propionibacteriaceae;  
Propionibacterium)**

**UCHIME score = 13.4277**

Default cutoff = 0.28



**Chimeras can be detected  
when the known reference sequences are given  
in the non-chimera database!**

# Non-chimera Reference Database in the EzBioCloud pipeline

- Manually curated high quality sequences of species with valid names
- >2,000 Genomospecies (new species based on genome sequence)
- >2,500 Trusted full-length sequences representing uncultured species (using Pacific Biosciences *ccs* sequencing)

# 16S sequence from human feces (V3V4 MiSeq)

>7194088

TGGGGAATATTGCACAATGGGCGAAAGCCTGATGCAGCAACGCCGCGTGAGTGAAGAAGTATCTC  
GGTATGTAAAGCTCTATCAGCAGGGAAGAAAATGACGGTACCTGACTAAGAAGCCCCGGCTAACTA  
CGTGCCAGCAGCCGCGGTAATACGTATGGGGCAAGCGTTATCCGGAATTACTGGGTGTAAAGGGAG  
CGTAGGCGGCATGGTAAGCCAGATGTGAAAGCCCGAAGCTTAAATTCGAGGATTGCATTTGGA  
ACTATCAAGCTAGAGTACAGGAGAGGAAAGCGGAATTCCTAGTGTAGCGGTGAAATGCGTAGATATTAG  
GAAGAACACCAGTGGCGAAGGCGGCTTTCTGGACTGAAACTGACGCTGAGGCTCGAAAGCGTGG  
GGAGCAAACA



Identify both fragments at <https://www.ezbiocloud.net/>



# Amplicon 7194088

- **Left fragment:** *Anaerotignum PAC001223\_s*  
Bacteria;Firmicutes;Clostridia;Clostridiales;Lachnospiraceae
- **Right fragment:** *DS990270\_s*  
Bacteria;Firmicutes;Clostridia;Clostridiales;Lachnospiraceae
- **UCHIME chimera score: 4.4643**



# Chimera ratios reported from published studies

	Hass <i>et al.</i> Genome Res. 2011.	Schloss <i>et al.</i> Plos One. 2011
Chimera detection rate	15-20%	5.5 – 15.9%
Mock community used	20 Bacterial + 1 Archaeal strains	20 Bacterial + 1 Arcaheal strains
PCR target region	V1/3, V3/5, V6/9	V1/3, V3/5, V6/9
Chimera detection method	ChimeraSlayer	Reference Database

\* Obtained from mock communities.



## Deep sequenced real data

- Human fecal sample of healthy Korean (male)
- V3V4 MiSeq sequencing
- 9,210,140 reads (single sample)
- Data browsable at
  - [https://www.ezbiocloud.net/chun\\_microbiome](https://www.ezbiocloud.net/chun_microbiome)

# Summary of the Chapter

- **Chimera filtering is an essential step in 16S rRNA-based microbiome profiling pipeline.**
- **Unfiltered chimeras will be recognized as new species/OTU; thereby species diversity can be over-estimated.**
- **The quality of chimera filtering depends on the quality of Non-chimera Reference Database (especially uncultured phylotypes/OTUs).**

