# Ancestry Informative Markers in the Skin Microbiome for Human Identification

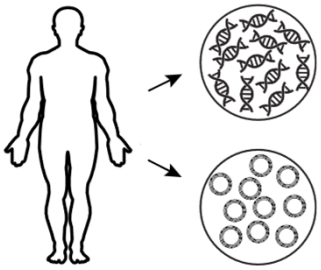
*Allison Sherier, August Woerner, Bruce Budowle*
*Center of Human Identification*
*University of North Texas of Health Science Center*

Promega | Powered for 32 years
ISHI
INTERNATIONAL SYMPOSIUM ON HUMAN IDENTIFICATION
ORLANDO, FL | SEPT. 13–17, 2021

1

## Current methods in human identification

hsc



Nuclear DNA
- High discriminatory power
- Touch samples – low quantity
- Degradation – low quality

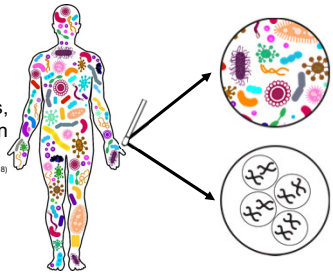
Mitochondrial DNA
- High quantity
- Low discriminatory power
- Maternally inherited

2

## Skin microbiome is another source of DNA

hsc

Skin swabbing only contained DNA equivalent to 4 diploid cells, but enough microbial information for accurate attribution
(Schmedes et al. 2018)



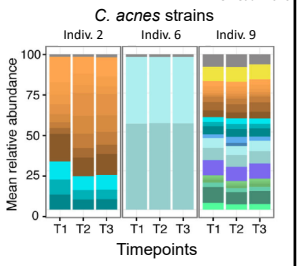
3

## Skin microbiome stable up to 3 years

hsc

Oh et al. 2016

- Shotgun sequence data
- Short time intervals (1-2 months)
  and long-time intervals (1-2 years)
- No substantial change
  in community composition




6

## Targeted genome sequencing panel

hsc

- Using Oh et al. 2016 data, Schmedes et al. 2018 identified stable microorganisms
  - *Cutibacterium acnes* was highly abundant
- 286 markers (ranging from family, genus, and species level resolution)

hidSkinplex

7

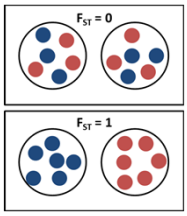
hsc

## Ancestry informative markers (AIMs)

8

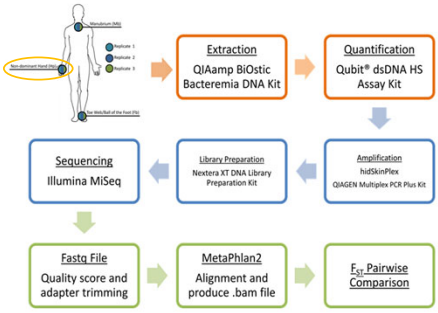
## Wright's index of fixation ($F_{ST}$)

- A measure of the difference in genetic structure between two populations



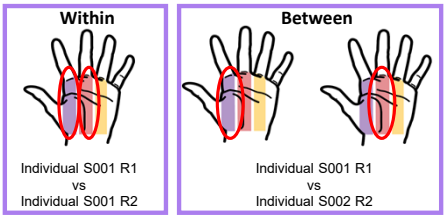

9

## Analysis workflow




10

$F_{ST}$ Comparisons between all sample pairs

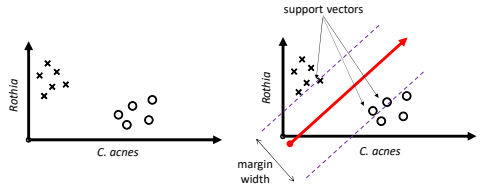



11

## What is support vector machine?

hsc

Classification method that finds a boundary to distinctly classify data points

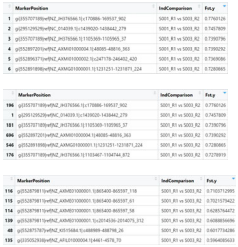



12

## SNPs for classification

hsc

- **Highest-ranking $F_{ST}$**
  Overall
  500, 1,000, or 2,000
- **Diversity**
  Per Marker
  5, 10, or 25
- **Common highest-ranking SNPs**
  Selected
  50 to 2,000

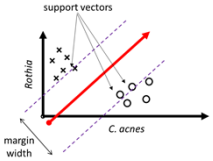


13

## Parameters

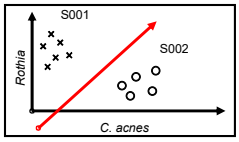
hsc

- Number of SNPs
- Minimum reads at each SNP
- Cost of SVM (helps determine optimal boundary)




14

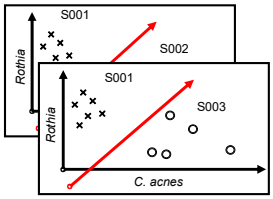
## SVM for hidSkinPlex




| Class | Confidence of Prediction | Vote |
|---|---|---|
| S001 | 0.82 | 1 |
| S002 | 0.18 | 0 |

15

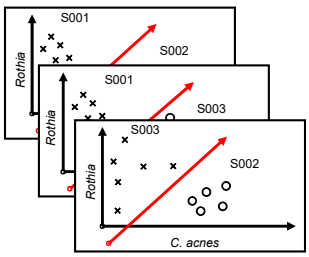
## SVM for hidSkinPlex




| Class | Confidence of Prediction | Vote |
|---|---|---|
| S001 | 0.82 | 1 |
| S002 | 0.18 | 0 |
| S001 | 0.75 | 1 |
| S003 | 0.25 | 0 |

16

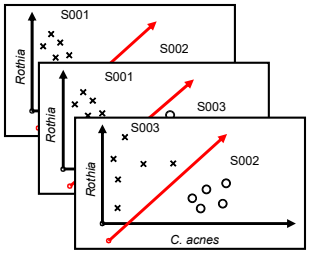
## SVM for hidSkinPlex




| Class | Confidence of Prediction | Vote |
|---|---|---|
| S001 | 0.82 | 1 |
| S002 | 0.18 | 0 |
| S001 | 0.75 | 1 |
| S003 | 0.25 | 0 |
| S002 | 0.37 | 0 |
| S003 | 0.63 | 1 |

17

## SVM for hidSkinPlex




| Class | Confidence of Prediction | Vote |
|---|---|---|
| S001 | 0.82 | 1 |
| S002 | 0.18 | 0 |
| S001 | 0.75 | 1 |
| S003 | 0.25 | 0 |
| S002 | 0.37 | 0 |
| S003 | 0.63 | 1 |

| Class | Confidence of Prediction | Vote |
|---|---|---|
| S001 | 0.78 | 2 |
| S003 | 0.63 | 1 |
| S002 | 0.00 | 0 |

18

## Accuracy of true positive

| Class | Confidence of Prediction | Vote |
|---|---|---|
| S001 | 0.79 | 25 |
| S002 | 0.67 | 21 |
| S003 | 0.00 | 0 |
| S004 | 0.67 | 20 |
| S005 | 0.69 | 15 |
| S006 | 0.69 | 12 |
| S007 | 0.73 | 24 |
| S008 | 0.65 | 16 |
| S009 | 0.71 | 3 |
| S010 | 0.61 | 4 |
| … | … | .. |

19

## Accuracy of true positive

| Class | Confidence of Prediction | Vote |
|---|---|---|
| S001 | 0.79 | 25 |
| S002 | 0.67 | 21 |
| S003 | 0.00 | 0 |
| S004 | 0.67 | 20 |
| S005 | 0.69 | 15 |
| S006 | 0.69 | 12 |
| S007 | 0.73 | 24 |
| S008 | 0.65 | 16 |
| S009 | 0.71 | 3 |
| S010 | 0.61 | 4 |
| … | … | .. |

Descending order by votes

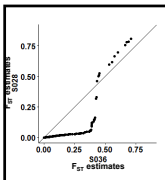
| Class | Confidence of Prediction | Vote |
|---|---|---|
| S001 | 0.79 | 25 |
| S007 | 0.73 | 24 |
| S028 | 0.76 | 23 |
| S038 | 0.70 | 22 |
| S002 | 0.67 | 21 |
| S004 | 0.67 | 20 |
| S029 | 0.65 | 19 |
| S033 | 0.69 | 18 |
| S035 | 0.69 | 17 |
| S008 | 0.65 | 16 |
| … | … | … |

20

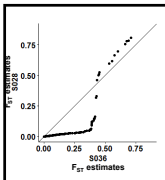
## Accuracy of true positive

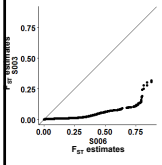
| Class | Classification | Result |
|---|---|---|
| … | … | … |
| S027 R2 | S027 | True |
| S027 R3 | S027 | True |
| S028 R1 | S028 | True |
| S028 R2 | S028 | True |
| S028 R3 | S036 | False |
| S029 R1 | S029 | True |
| S029 R2 | S028 | False |
| S029 R3 | S029 | True |
| S030 R1 | S030 | True |
| … | … | .. |

21

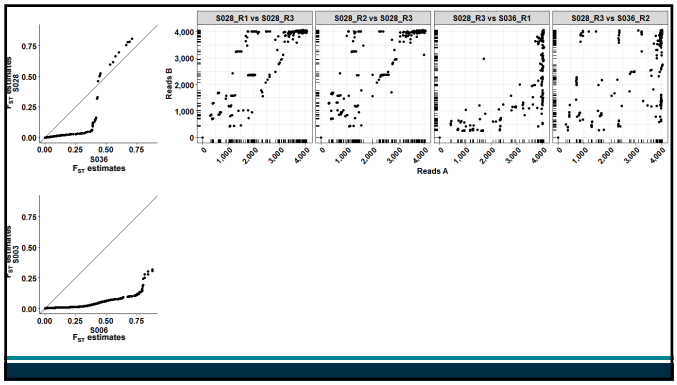



24

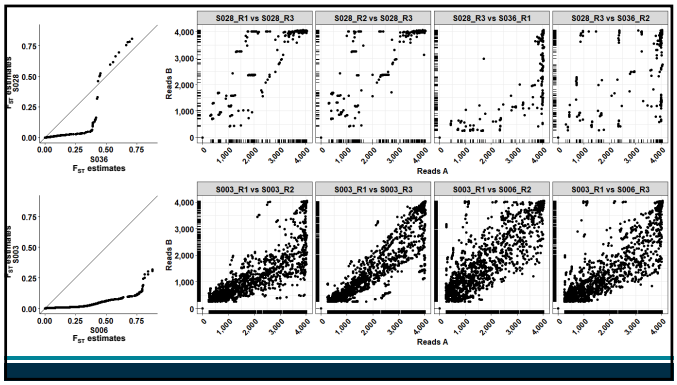



25




26



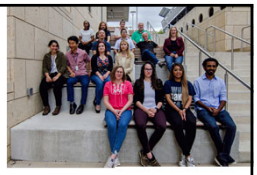

27

## Acknowledgements

Bruce Budowle
August Woerner
Jonathan King
Sarah Schmedes
Nicole Novroski
Rachel Kieser
Frank Wendt
Utpal Smart
Ben Cyrsup
Sammed Mandape

Research and Development Unit
Center for Human Identification

Funded in part by National Institute of Justice:
2015-NE-BX-K006 (PI: Dr. Bruce Budowle)
2020-R2-CX-0046 (PI: Allison Sherier)



29