The human milk microbiome
healthy breasts, mothers, and babies

James A. Foster
22 September 2010
The human milk microbiome
healthy breasts, mothers, and babies

James A. Foster
22 September 2010
Who cares about breast milk?

introduction ➰ who cares?

Who cares about breast milk?

introduction ➔ who cares?

춘Approximately half of all existing adults
  • *Mastitis*: choose between intense pain, starving your infant, or using less healthy “formula” (often with dirty water)
Who cares about breast milk?

- Approximately half of all existing adults
  - *Mastitis*: choose between intense pain, starving your infant, or using less healthy “formula” (often with dirty water)

- Nursing infants (future adults)
  - *Intestinal disease* is the number one infant killer in the world
  - which is “better”: breast milk or formula, given that clean water is a luxury in most places
Key questions

✧ Is there a “milk microbiome”?
✧ What is the ecology of the milk microbiome?
  • Which bacteria are present? (who?)
  • How complex is the milk microbiome?
Key questions

Is there a “milk microbiome”?

What is the ecology of the milk microbiome?
- Which bacteria are present? (who?)
- How complex is the milk microbiome?
- Where do(es) the milk microbiome(s) come from?
- Is a woman’s microbiome stable over time?
- What are the bacteria doing?

Are there “healthy” or “unhealthy” milk microbiomes?
- Maternal disease (e.g. mastitis)
- Infant disease (e.g. intestinal disease, sepsis, premies)
- Maternal and infant health (e.g. prophylaxis, positive benefit)
Key questions

Is there a “milk microbiome”?

What is the ecology of the milk microbiome?
  - Which bacteria are present? (who?)
  - How complex is the milk microbiome?
    - Where do(es) the milk microbiome(s) come from?
    - Is a woman’s microbiome stable over time?
    - What are the bacteria doing?

Are there “healthy” or “unhealthy” milk microbiomes?
  - Maternal disease (e.g. mastitis)
  - Infant disease (e.g. intestinal disease, sepsis, premies)
  - Maternal and infant health (e.g. prophylaxis, positive benefit)
What is “known”

introduction ☛ is there milk microbiome?
What is “known”

—is there milk microbiome?

Two dogmas

• Mom’s: “Duh, of course there are bacteria in milk”
• Medical: “of course not, healthy milk is ‘sterile’ ”
  - translation: “if there aren’t any pathogens, who cares?”
  - evidence: milk on plates doesn’t produce (many) colonies, and what we find looks like it came from the skin [J. Appl. Microbiol. 95, 471-478 (2003)]
What is “known”

Two dogmas

• Mom’s: “Duh, of course there are bacteria in milk”
• Medical: “of course not, healthy milk is ‘sterile’ ”
  - translation: “if there aren’t any pathogens, who cares?”
  - evidence: milk on plates doesn’t produce (many) colonies, and what we find looks like it came from the skin [J. Appl. Microbiol. 95, 471-478 (2003)]

Our study

• 16 women, 3 time points over 4 weeks, collected aseptically
• barcoded pyrosequencing of V1-V3 region of 16S rRNA gene: over 230K sequences, average 4000 reads per sample, approximately 240bp each
• bioinformatics: RDP classifier, mothur, R
## Data reduction: quality control

<table>
<thead>
<tr>
<th>QC Step</th>
<th>#/reads</th>
</tr>
</thead>
<tbody>
<tr>
<td>Raw data</td>
<td>448,051 (233bp avg)</td>
</tr>
<tr>
<td>Remove reads/w low quality scores, ambiguities, homopolymers</td>
<td>435,036</td>
</tr>
<tr>
<td>Align, remove unalignable characters</td>
<td>301,876 (40,005 unique)</td>
</tr>
<tr>
<td>Precluster, remove chimeras</td>
<td>271,180 (16,327 unique, 60% of original)</td>
</tr>
</tbody>
</table>
Data reduction: quality control

<table>
<thead>
<tr>
<th>QC Step</th>
<th>#/reads</th>
</tr>
</thead>
<tbody>
<tr>
<td>Raw data</td>
<td>448,051 (233bp avg)</td>
</tr>
<tr>
<td>Remove reads/with low quality scores, ambiguities, homopolymers</td>
<td>435,036</td>
</tr>
<tr>
<td>Align, remove unalignable characters</td>
<td>301,876 (40,005 unique)</td>
</tr>
<tr>
<td>Precluster, remove chimeras</td>
<td>271,180 (16,327 unique, 60% of original)</td>
</tr>
</tbody>
</table>

Length of reads for breastmilk microbiome

- Median = 255bp
- SD = 65.674

Tuesday, October 5, 2010
## Data reduction: quality control

<table>
<thead>
<tr>
<th>QC Step</th>
<th>#/reads</th>
</tr>
</thead>
<tbody>
<tr>
<td>Raw data</td>
<td>448,051 (233bp avg)</td>
</tr>
<tr>
<td>Remove reads/w low quality scores, ambiguities, homopolymers</td>
<td>435,036</td>
</tr>
<tr>
<td>Align, remove unalignable characters</td>
<td>301,876 (40,005 unique)</td>
</tr>
<tr>
<td>Precluster, remove chimeras</td>
<td>271,180 (16,327 unique, 60% of original)</td>
</tr>
</tbody>
</table>

### Length of reads for breastmilk microbiome

- Median: 255bp
- Standard deviation: 65.674

- Median: 191bp
- Standard deviation: 7.1177

---

rRNA amplicon bioinformatics

methods ➫ 16S fingerprinting
rRNA amplicon bioinformatics

methods ➔ 16S fingerprinting

**Guiding principle:**
when in doubt, throw it out

16,327 unique sequences representing 301,867 high quality reads (from 448,051)
rRNA amplicon bioinformatics

Guiding principle: when in doubt, throw it out

<table>
<thead>
<tr>
<th>Bioinformatics analysis</th>
<th>Sanity Checks</th>
<th>Community analysis</th>
</tr>
</thead>
<tbody>
<tr>
<td>Initial QC</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Remove short, low quality, ambiguous reads</td>
<td>Check number, length of pruned &amp; retained reads, homopolymers</td>
<td>Define treatments</td>
</tr>
<tr>
<td>Who's there?</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Classify sequences</td>
<td>Biologically likely?</td>
<td>Estimate richness, evenness</td>
</tr>
<tr>
<td>Ecological analysis</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Align sequences: use secondary structure</td>
<td>Are number, lengths, quality reasonable?</td>
<td>Estimate rarefaction, richness, evenness</td>
</tr>
<tr>
<td>Trim ends, filter characters</td>
<td>OTU names match above?</td>
<td></td>
</tr>
<tr>
<td>Determine OTUs: compute similarity, cluster</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Select/classify OTU representatives</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

16,327 unique sequences representing 301,867 high quality reads (from 448,051)

Microbiomes are complex

Estimated and observed richness

ACE richness estimate with 95% confidence intervals (open circles) and observed richness (vertical bars).

Mothers are similar

results ➞ core microbiome

Genera are consistent within each woman over time, and all women share a “core” of four major genera.

Mothers are different

Comparison of microbial community structures

Bray-Curtis measure of community similarity: women's communities are similar over time but differ from other women.
“similar but different”

Proportion and Relative Abundance of Persistant OTUs

- %OTUs persistant across three timepoints
- Relative abundance of persistant OTUs

Subject Identification

S1 S2 S3 S4 S5 S6 S7 S8 S9 S10 S11 S12 S13 S14 S15 S16
Conclusions

Conclusions ➔ this study
Conclusions

There is a (rich) milk microbiome that differs significantly from woman to woman.
Conclusions

- There is a (rich) milk microbiome that differs significantly from woman to woman
- There is a “core microbiome” that all women share, comprising a few very abundant genera
Conclusions

✦ There is a (rich) milk microbiome that differs significantly from woman to woman
✦ There is a “core microbiome” that all women share, comprising a few very abundant genera
✦ Mothers’ microbiome differ significantly, due to very rich and diverse collection of scarce bacteria
Next steps

discussion ➔ next steps
Next steps

- Are these results representative? Need more samples from more mothers over longer periods of time.

- Refined resolution of ecological questions
  - “drill down” to strain level when possible
  - Where do the milk microbiomes come from? Succession?
  - Is a woman’s microbiome stable? Are there cycles? Are there fixed points?
  - What are the bacteria doing?

- Are there “healthy” or “unhealthy” milk microbiomes? What can the microbiomes tell us about disease diagnosis and treatment and preventative care?
  - Maternal and infant health (e.g. prophylaxis, positive benefit)
  - Maternal disease (mastitis)
  - Infant disease (e.g. intestinal disease, sepsis, preemies)

- Develop better bioinformatics
Acknowledgements

Human Microbiome

- Katherine Hunt
- Mark McGuire
- Michelle McGuire
- Daniel Beck
- Larry Forney
- Ursel Schüette
- Zaid Abdo
- Wade Copeland
- (many more!)

IBEST Facilities

- Matt Settles
- Suresh Iyer
- Dan
- Rob Lyon
- Trent Nelson
- Colby Blair

Funding

- COBRE (Forney) NIH P20 RR16448
- INBRE (Bohach) NIH P20 RR016454
- NSF “Evolution in Action” STC DBI 0939454
Let the conversations begin!
ACE richness estimation

Essentially: estimated richness = observed richness + unobserved richness estimated from proportion of rare to common species ("the tail")

\[
S_{ACE} = \begin{cases} S_{abund} + \frac{S_{rare}}{C_{ACE}} + \frac{n_1}{C_{ACE}} \hat{\gamma}_{ACE}^2, & \text{for } \hat{\gamma}_{ACE} < 0.80 \\ S_{abund} + \frac{S_{rare}}{C_{ACE}} + \frac{n_1}{C_{ACE}} \hat{\gamma}_{ACE}^2, & \text{for } \hat{\gamma}_{ACE} \geq 0.80 \end{cases}
\]

\[
N_{\text{rare}} = \sum_{i=1}^{\text{abund}} in_i
\]

\[
C_{ACE} = 1 - \frac{n_1}{N_{\text{rare}}}
\]

\[
\hat{\gamma}_{ACE} = \max \left\{ \hat{\gamma}_{ACE}^2 \left\{ 1 + \frac{N_{\text{rare}}(1 - C_{ACE}) \sum_{i=1}^{\text{abund}} i(i-1)n_i}{N_{\text{rare}}(N_{\text{rare}} - C_{ACE})} \right\}, 0 \right\}
\]

\[
\hat{\gamma}_{ACE}^2 = \max \left[ \frac{S_{rare} \sum_{i=1}^{\text{abund}} i(i-1)n_i}{C_{ACE} N_{\text{rare}}(N_{\text{rare}} - 1) - 1}, 0 \right]
\]

\[ni = \text{The number of OTUs with } i \text{ individuals}\]

\[Srare = \text{The number of OTUs with 10 or fewer individuals}\]

\[Sabund = \text{The number of OTUs with more than 10 individuals}\]

\[abund = \text{the threshold to be considered an 'abundant' OTU (10)}\]

(equations from Mothur documentation)
Bray-Curtis structural similarity

Bray-Curtis community comparison

\[ C_{Bray-Curtis} = 2 \frac{\sum \min(S_{A,i}, S_{B,i})}{\sum S_{A,i} + \sum S_{B,i}} \]

where,

\[ S_{A,i} = \text{the number of individuals in the } i\text{th OTU of community A} \]

\[ S_{B,i} = \text{the number of individuals in the } i\text{th OTU of community B} \]

Essentially: ratios of abundances in correspondingly ranked OTUs

(equations from Mothur documentation)
Data reduction: Quality control

The original 454 run produced 448,051 reads

Sequences were assigned to a group for each specific sample—based on their barcodes—before being trimmed to remove barcodes and primers. Processing removed low-quality reads and those with ambiguous bases, large homopolymers, or those that did not include correct barcode or primer sequences. This left 435,036 reads (%97 of the original data set).

For ease of computation during alignment and distance matrix construction, all redundant sequences were removed from the pipeline and placed in a file for use in downstream processing, leaving 86,106 unique sequences.

Sequences were then aligned to Mothur’s SILVA database and processed to produce a set of sequences that aligned over the same region (V2) of the 16S gene, and removing those that did not. This left 64,391 unique sequences representing 301,867 total reads (%67 of the original data set).

Aligned sequences were then trimmed to remove columns that contained only gaps or a “.” character. This generated new redundant sequences which were then removed and placed in a file for later use in downstream processing to leave 40,005 unique sequences.

To reduce the effects of sequencing error on richness estimates, a pre-cluster function that has been shown to be effective (Huse 2010) was then employed, which combines sequences that differ by one character. This reduced the number of unique sequences to 21,446.

As a final quality control measure the Chimera Slayer program was then used to remove potential chimeras from the data set.

• The final result of quality control produced 16,327 unique, high quality sequences representing a total of 271,180 total sequences (60% of the original data set) for use in downstream analysis.