Applications of Next Generation Sequencing in Metagenomics Studies

Francesca Rizzo, PhD
Genomix4life
Laboratory of Molecular Medicine and Genomics
Department of Medicine and Surgery
University of Salerno
francesca.rizzo@genomix4life.com
frizzo@unisa.it
Introduction

<table>
<thead>
<tr>
<th>Genome = Parts list of a single genome</th>
</tr>
</thead>
<tbody>
<tr>
<td>- Human genome: 3 Gbp, 24,000 ORFs</td>
</tr>
<tr>
<td>- Bacterial Genome: 3-5 Mbp, 200-4000 ORFs</td>
</tr>
<tr>
<td>- Viral Genome: 0.1-1Mbp, 10-100 ORFs</td>
</tr>
<tr>
<td>- e.g.: Garden soil</td>
</tr>
<tr>
<td>- 50 archeal species: 1.5x10^8 bp</td>
</tr>
<tr>
<td>- 1000 bacterial species: 5.0x10^9 bp</td>
</tr>
<tr>
<td>- 200 fungal species: 1.0x10^{11} bp</td>
</tr>
<tr>
<td>- 50 invertebrate species: 5.0x10^{10} bp</td>
</tr>
<tr>
<td>- 5000 viral genomes: 5.0x10^7 bp</td>
</tr>
</tbody>
</table>

**Meta**genome = Parts list of the community

“...functional analysis of the collective genomes of soil microflora, which we term the **metagenome** of the soil.”
- J. Handelsman et al., 1998
"the application of modern genomics techniques to the study of communities of microbial organisms directly in their natural environments, bypassing the need for isolation and lab cultivation of individual species"

"the application of modern genomics techniques to the study of communities of microbial organisms directly in their natural environments, bypassing the need for isolation and lab cultivation of individual species"

Metagenomics

Who is there?

What are they doing?
NGS methods

Marker Genes
(Amplification sequencing)
- Extract DNA
- Amplify with targeted primers
- Diversity analysis

Matagenomics
(Shotgun sequencing)
- Extract DNA
- Sequence random fragments
- Diversity, function analysis

Metatranscriptomics
(RNA-sequencing)
- Extract RNA
- Sequence cDNA
- Gene expression, function
NGS methods

**Marker Genes**
(Amplicon sequencing)

1. Extract DNA
2. Amplify with targeted primers
3. Diversity analysis

**Matagenomics**
(Shotgun sequencing)

1. Extract DNA
2. Sequence random fragments
3. Diversity, function analysis

**Metatranscriptomics**
(RNA-sequencing)

1. Extract RNA
2. Sequence cDNA
3. Gene expression, function
Marker Genes (Amplicon sequencing)

16S rRNA Genes

- V1
- V2
- V3
- V4
- V5
- V6
- V7
- V8
- V9

Internal transcribed spacer (ITS)

- Bacteria
- Fungi

Alternative gene markers

- **recA**: Response to DNA Stress in Bacteria
- **18S rRNA**: 18S ribosomal RNA
- **rpoB**: RNA polymerase β subunit
- **hsp70**: heat shock protein 70
Typical workflow

Marker Genes (Amplicon sequencing)

Extract DNA → Library preparation → DNA

Environmental sample

- Forward primer overhang adapter
- V3 region specific primer
- PCR
- Adapter 1 16S amplicon Adapter 2
- Nextera® XT indexes
- Adapter 1 16S amplicon Adapter 2
- PCR
- Normalization and pooling
- MiSeq sequencing
Typical workflow

Marker Genes (Amplicon sequencing)

Environmental sample → Extract DNA → Library preparation → DNA → PCR → Normalization and pooling → MiSeq sequencing
**Typical workflow**

**Marker Genes (Amplicon sequencing)**

- **Extract DNA**
- **Library preparation**
- **Sequencing**
- **Analysis**

**Environmental sample**

**DNA**

**Platforms:**
- Illumina MiSeq (2X150 bp, 2X250 bp, 2X300 bp)

**Depth of sampling:**
- 50,000-100,000 reads/sample for simple communities
- 200,000-500,000 reads/sample for complex metagenomes
Typical workflow

Marker Genes (Amplicon sequencing)

Barcodes are used to assign each sequence to the sample it come from, dropping low quality reads.

Adapter are trimmed and multiple sequence alignment based on reference sequences is built.

Phylogenetic tree is built using representative OTU.

Related sequences are grouped into OTUs and taxonomy is assigned.

Use community clustering techniques to relates samples to one another.

Who is there?

Quantitative assessment of the community

OTU (Operational Taxonomic Unit)
Marker Genes (Amplicon sequencing)

ADVANTAGE:
- Fast survey of large community
- Reduced cost
- Mature set of tools and statistics for analysis
- Good for first round surveys

DISADVANTAGE:
- Lack of tools for processing ITS/fungal microbiome data set
- Amplification bias affect accuracy and replication
- 16S or ITS may not differentiate between similar strains
- Different target regions in 16S gene and ITS region can give different results/resolution
- Chloroplast 16S sequences can get amplified in plant metagenomes
NGS methods

Marker Genes (Amplicon sequencing)

- Extract DNA
- Amplify with targeted primers
- Diversity analysis

Matagenomics (Shotgun sequencing)

- Extract DNA
- Sequence random fragments
- Diversity, function analysis

Metatranscriptomics (RNA-sequencing)

- Extract RNA
- Sequence cDNA
- Gene expression, function
Matagenomics (Shotgun sequencing)

Typical workflow

1. **Extract DNA**
2. **DNA fragmentation**
3. **End repair**
4. **3’ End adenylation**
5. **Adapter ligation**
6. **PCR**
7. **Library preparation**
Matagenomics (Shotgun sequencing)

**Typical workflow**

**Environmental sample**

1. Extract DNA
2. DNA fragmentation
3. Library preparation
   - End repair
   - 3' End adenylation
   - Adapter ligation
4. PCR
5. Sequencing
6. Analysis
   - Number of species?
   - Relative abundance of the species?
   - Known or unknown genomes?

**PLATFORM:**
- Illumina NextSeq (2X150 bp), HiSeq X (2X125 bp)

**DEPTH OF SAMPLING:**
- 50,000,000-500,000,000 reads/sample
Typical workflow

Matagenomics (Shotgun sequencing)

Barcodes are used to assign each sequence to the sample it come from, dropping low quality reads.

Who is there?
What are they doing?
What can they do?

Metagenomes alignment to reference genome and/or de novo assembling.

Taxonomic classification
Function analysis
Gene prediction and/or annotation
Gene/genome structure

Genome A
Genome B
Genome C
Genome D
Unknown

Protein family
Gene ontology
Metabolic pathways
NGS methods

**Marker Genes**
(Amplicon sequencing)

- Extract DNA
- Amplify with targeted primers
- Diversity analysis

**Matagenomics**
(Shotgun sequencing)

- Extract DNA
- Sequence random fragments
- Diversity, function analysis

**Metatranscriptomics**
(RNA-sequencing)

- Extract RNA
- Sequence cDNA
- Gene expression, function
Introduction

Genome = Parts list of a single genome

Human genome:
3 Gbp, 24,000 ORFs

Bacterial Genome:
3-5 Mbp, 200-4000 ORFs

Viral Genome:
0.1-1Mbp, 10-100 ORFs

e.g.: Garden soil

- 50 archeal species: 1.5x10^8 bp
- 1000 bacterial species: 5.0x10^9 bp
- 200 fungal species: 1.0x10^11 bp
- 50 invertebrate species: 5.0x10^10 bp
- 5000 viral genomes: 5.0x10^7 bp

Meta-genome = Parts list of the community

160 Gb DNA
8 Gb RNA
Metatranscriptomics (RNA-sequencing)

Typical workflow

Environmental sample → Extract RNA → Total RNA → rRNA depletion → Optional → Library preparation
Typical workflow

Metatranscriptomics (RNA-sequencing)

Extract RNA

Environmental sample

RNA fragmented and primed

First strand cDNA synthesis

Second strand cDNA synthesis

3’ End adenylation

Adapter ligation

PCR

rRNA depletion (Optional)

Library preparation
Typical workflow

Metatranscriptomics (RNA-sequencing)

Environmental sample

Extract RNA

rRNA depletion (Optional)

Library preparation

Sequencing

Analysis

PLATFORM:
Illumina NextSeq (2X150 bp), Hiseq X (2X125bp)

Read 1: 150 bp
Index 1

Index 2

Read 2: 150 bp

DEPTH OF SAMPLING:
20,000,000-500,000,000 reads/sample

Number of species?
Relative abundance of the species?
Known or unknown genomes?
Levels of gene expression?
Barcodes are used to assign each sequence to the sample it came from, dropping low quality reads.

Alignment to reference transcriptome and/or de novo assembling.

Who is there?
What are they doing?

Protein family profile
Function analysis
Gene ontology abundance
Pathways inference and enrichment analysis
Metabolic pathways reconstruction
Taxonomic classification
New molecules identification

Typical workflow
Metatranscriptomics (RNA-sequencing)
Who is there

What are they doing?
WHY?

Metagenomics

Discovery of:

- the diversity of life
- what is a genome/species
- the interplay between human and microbes
- how do microbial communities work and how stable are
- novel metabolic pathways
- novel natural products
- new antibiotic
- new molecules with new functions
- new enzymes and bioactive molecules
NEXT GENERATION SEQUENCING

- DNA
  Whole genome & Exome sequencing targeted resequencing - de novo sequencing
- RNA
  RNA-Seq - smallRNA-Seq - RIP-Seq - CAGE
- Epigenomics and Gene regulation
  ChIP-Seq - BS-Seq - targeted BS-Seq - methyl-Seq
- Metagenomics and metatranscriptomics
  Amplicon-Seq - Shotgun-Seq
- Single cell sequencing

MICROARRAY

- Genotyping
  Infinium HD - Vera Code Golden Gate
- Epigenetics
  Human Methylation EPIC BeadChip
- Nutrigenomics
Applications of Next Generation Sequencing in Metagenomics Studies

Francesca Rizzo, PhD
Genomix4life
Laboratory of Molecular Medicine and Genomics
Department of Medicine and Surgery
University of Salerno
francesca.rizzo@genomix4life.com
frizzo@unisa.it