



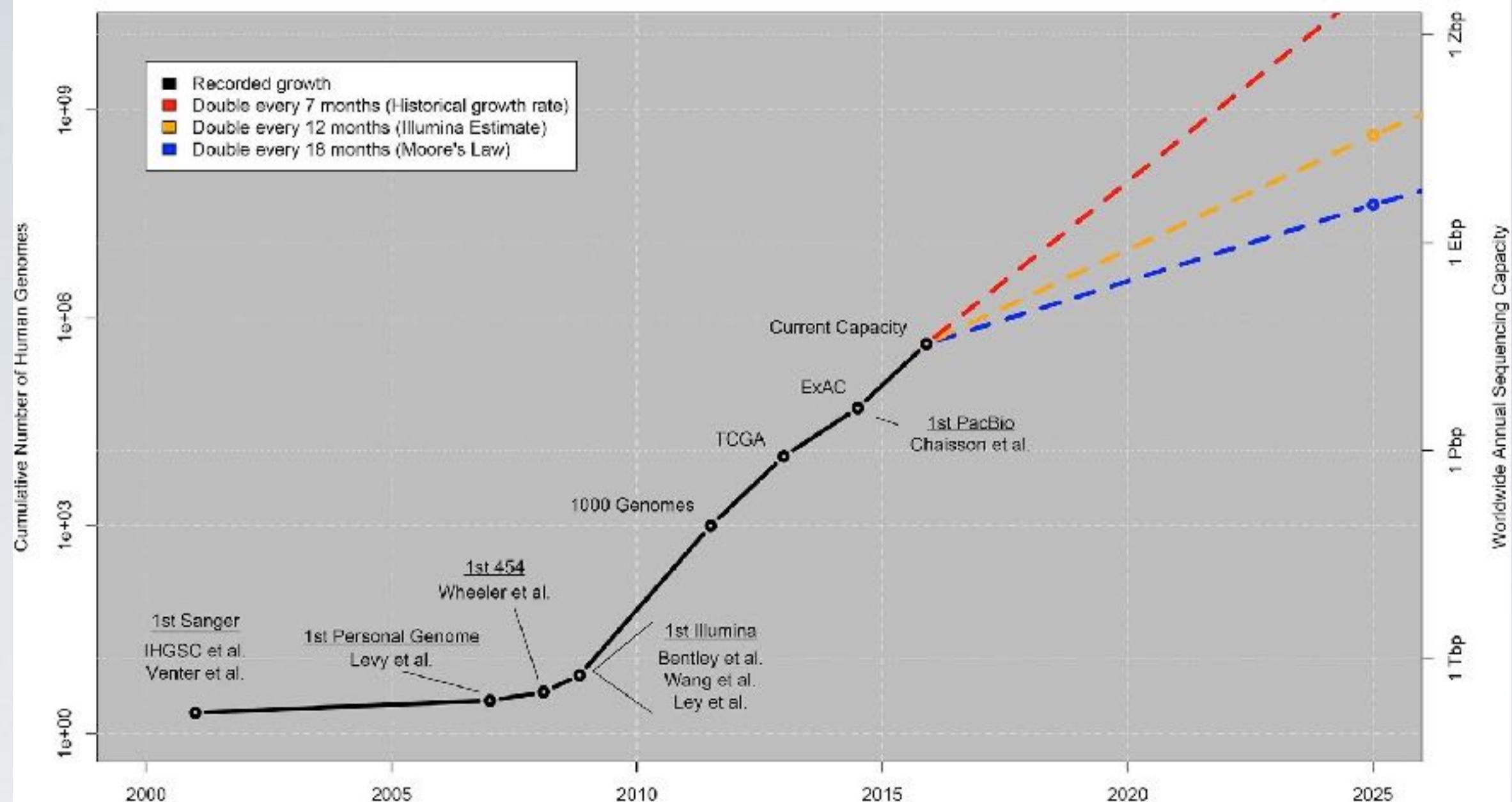
# BIONODE.IO

#MozFest 2016

[github.com/bionode](https://github.com/bionode)

[gitter.im/bionode/bionode](https://gitter.im/bionode/bionode)





# Growth of DNA Sequencing



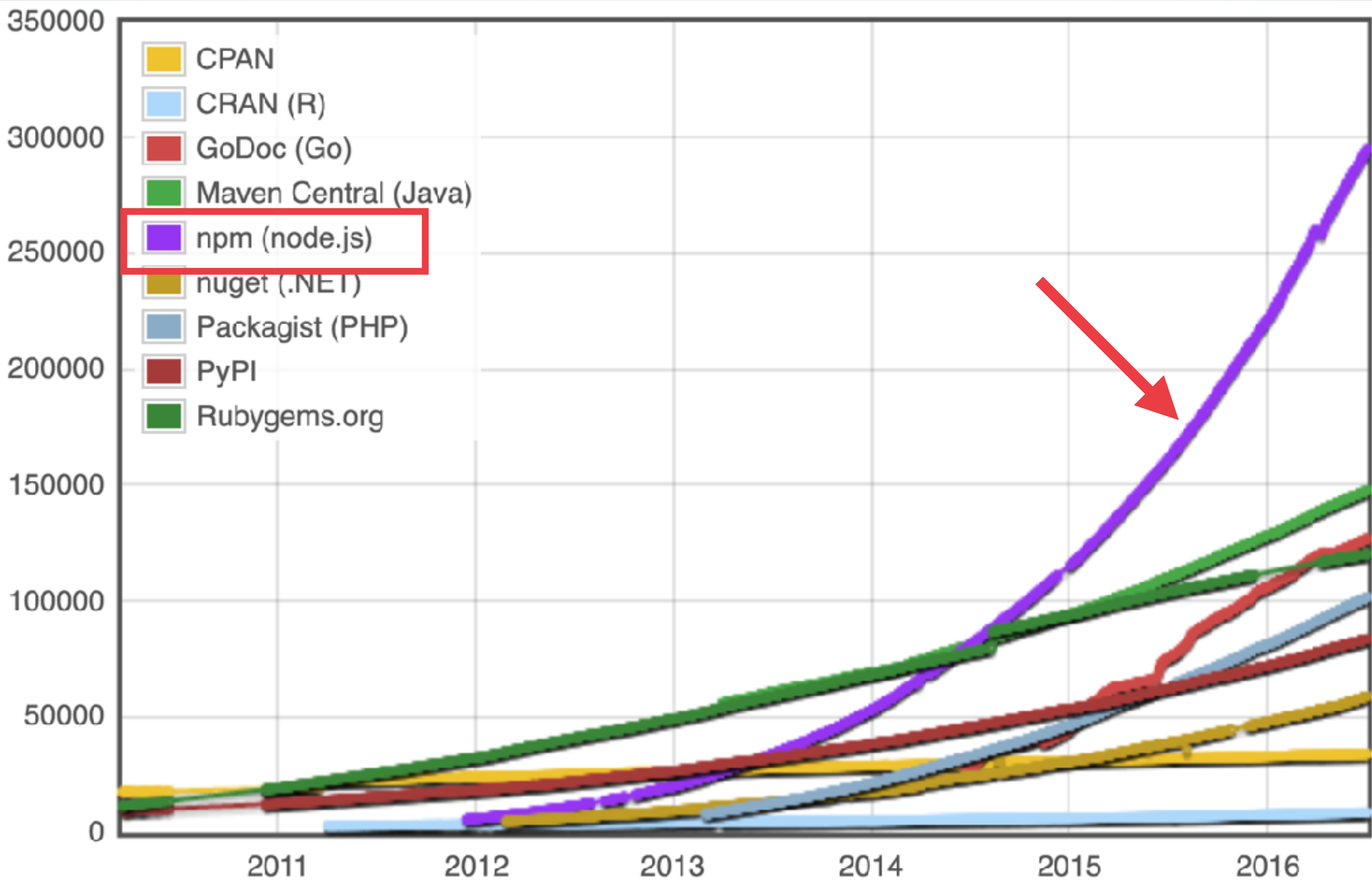
Data Phase	Astronomy	Twitter	Year YouTube	Genomics
Acquisition	25 zetta-bytes/year	0.5–15 billion tweets/year	500–900 million hours/year	1 zetta-bases/year
Storage	1 EB/year	1–17 PB/year	1–2 EB/year	2–40 EB/year
Analysis	In situ data reduction	Topic and sentiment mining	Limited requirements	Heterogeneous data and analysis
	Real-time processing	Metadata analysis		Variant calling, ~2 trillion central processing unit (CPU) hours
	Massive volumes			All-pairs genome alignments, ~10,000 trillion CPU hours
Distribution	Dedicated lines from antennae to server (600 TB/s)	Small units of distribution	Major component of modern user's bandwidth (10 MB/s)	Many small (10 MB/s) and fewer massive (10 TB/s) data movement

# WHAT

Modular and universal bioinformatics

-  Each tool tries to do one thing well
-  Provides highly reusable code and tools
-  Scales by using Streams
-  Runs everywhere

## Modules count




# HOW

## Using Node.js

 Highly modular

 Very open community on GitHub

 Provides native implementation of Streams

 Run same JavaScript code on browser or CLI

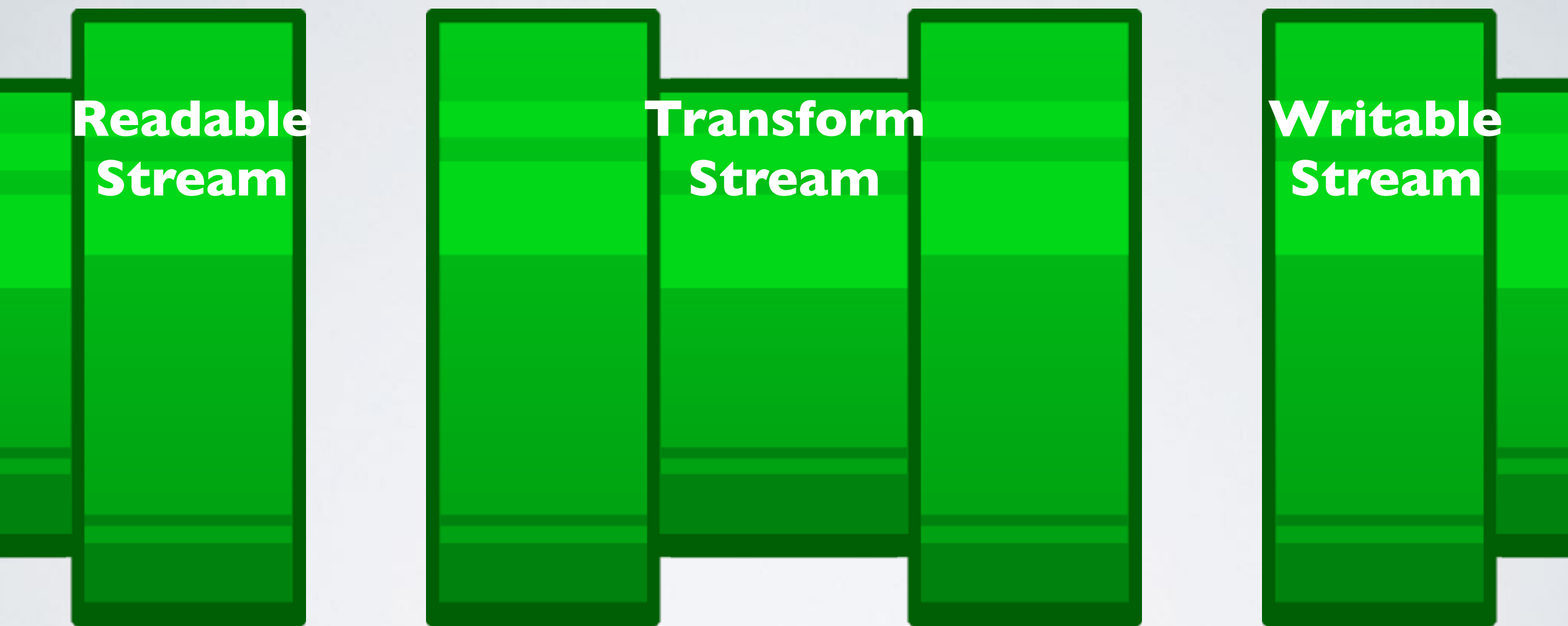


During my PhD at

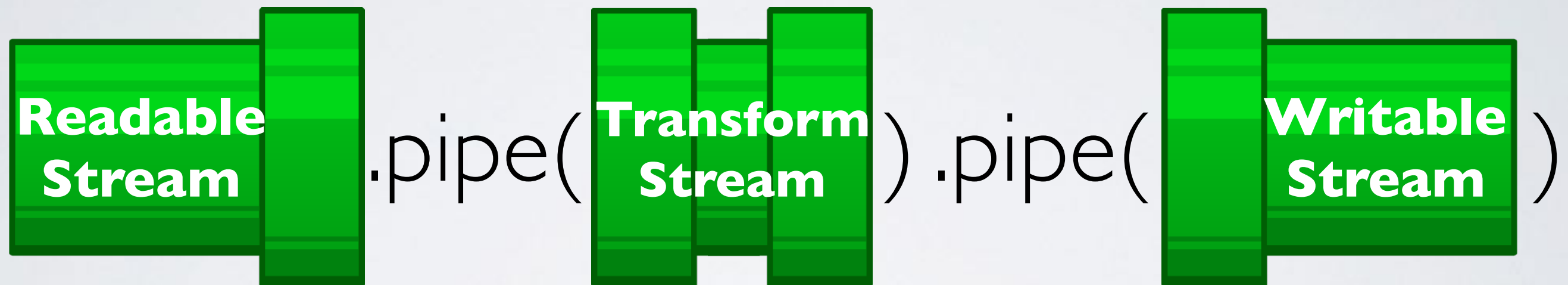
Compare genetic diversity of  
social vs solitary species



- Involved in biological web projects that need JS
- Had to find and get TB of data online
- Had to build complicated bioinformatic pipelines



# PROCESS DATA IN CHUNKS



`fs.createReadStream(file)`

`request(url)`

`process.stdin()`

`JSONStream.parse()`

`filterFunction()`

`multithreadAnalysis()`

`fs.createWriteStream(file)`

`process.stdout()`



**Readable  
Stream**

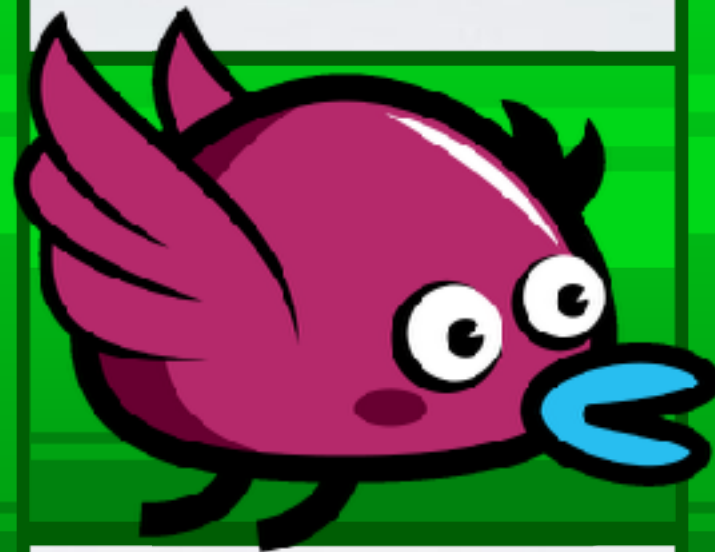
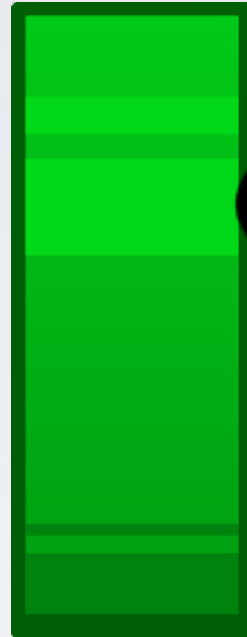
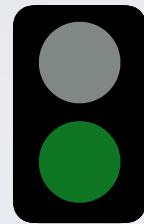


**Writable  
Stream**

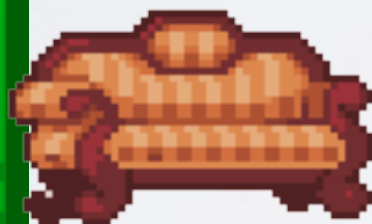


**this.push(data)**  
**FALSE**

**Readable  
Stream**



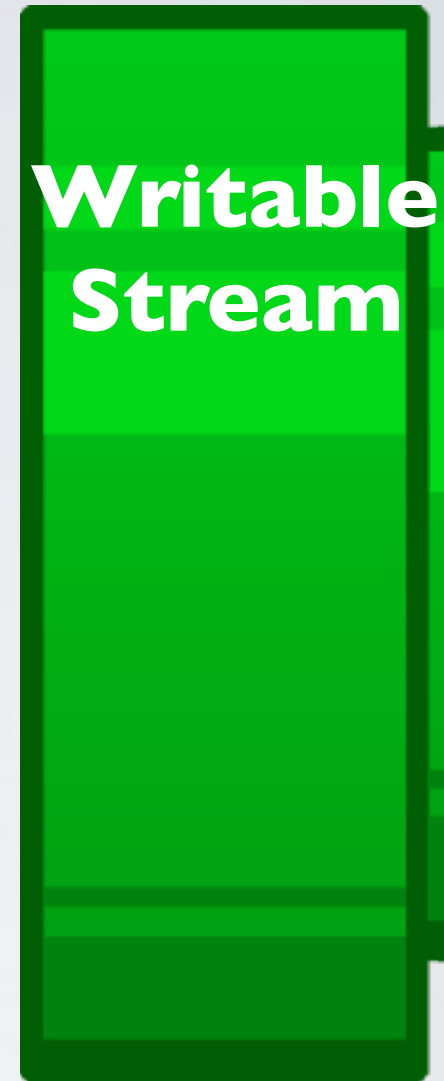
**callback()**



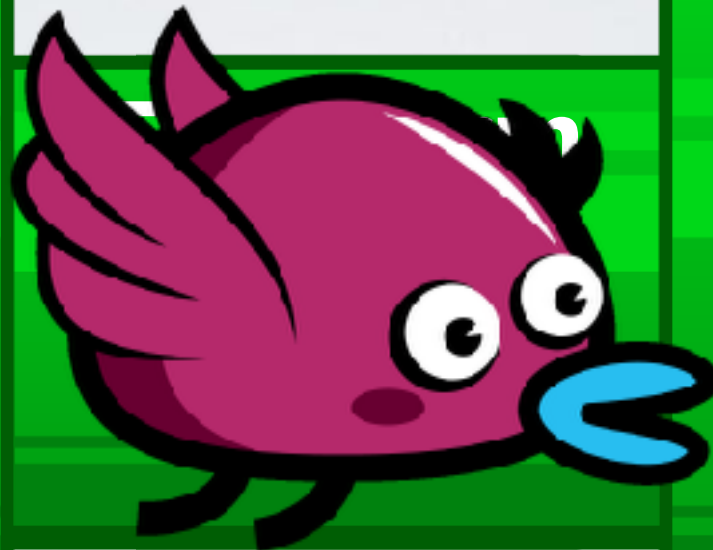
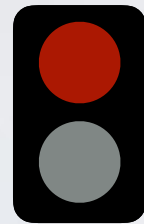
**Buffer**

**this.push(data)**  
**TRUE**

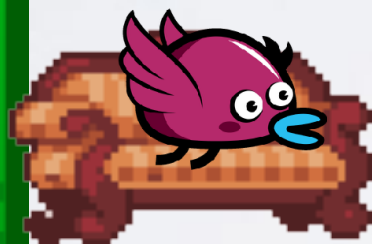
**Writable  
Stream**



**Readable  
Stream**



**Writable  
Stream**

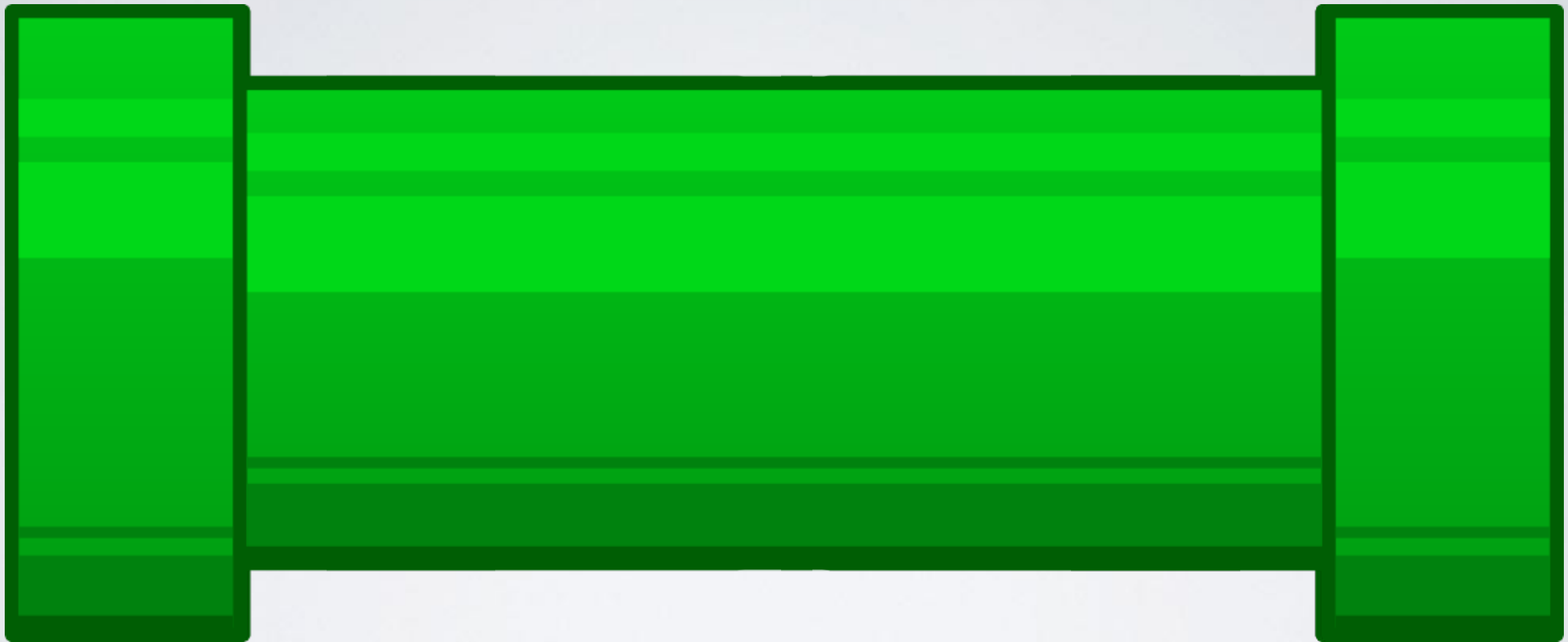


**Buffer**

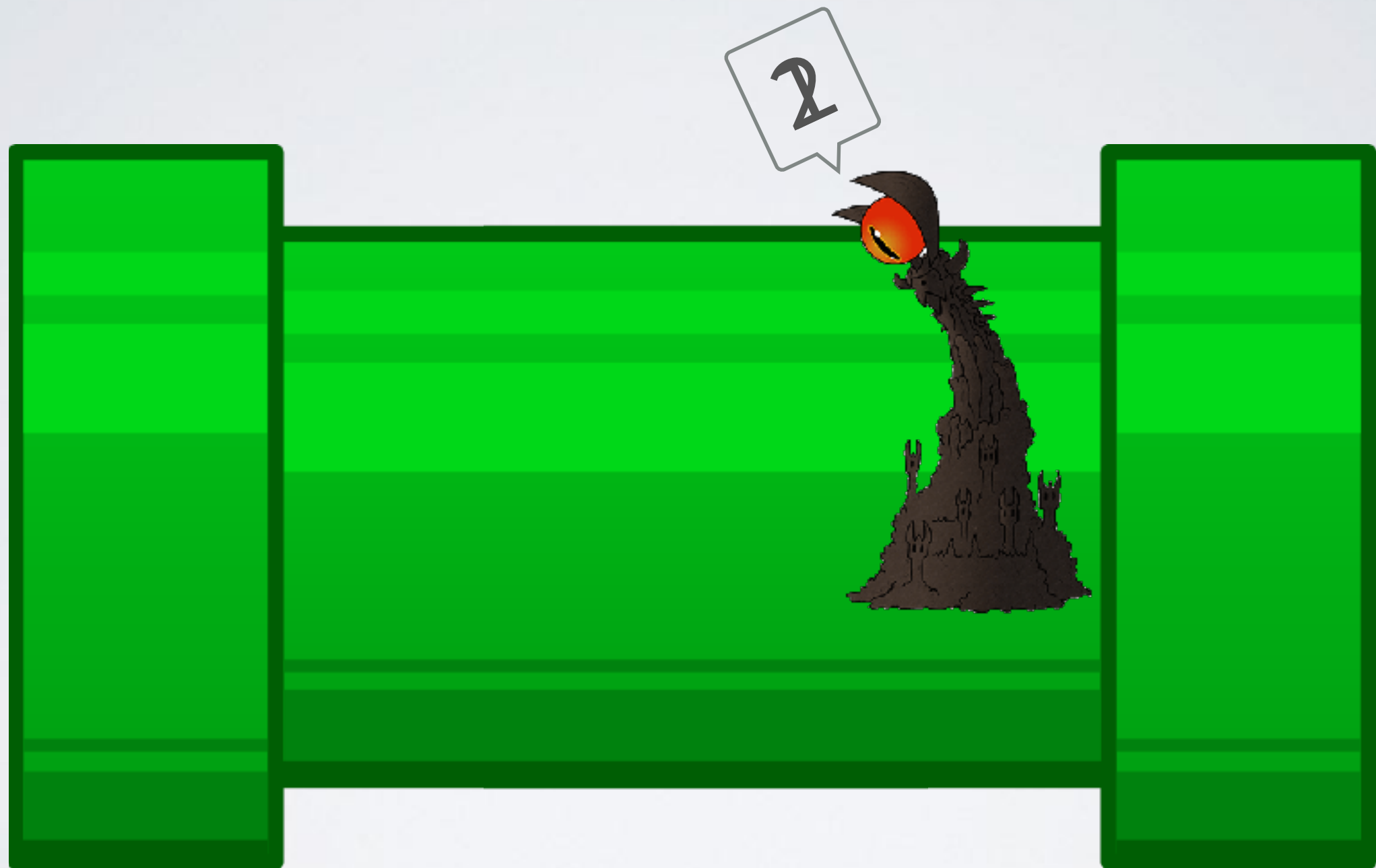
**callback()**

**this.push(data)  
TRUE**

# DUPLEX STREAM



# PASSTHROUGH STREAM

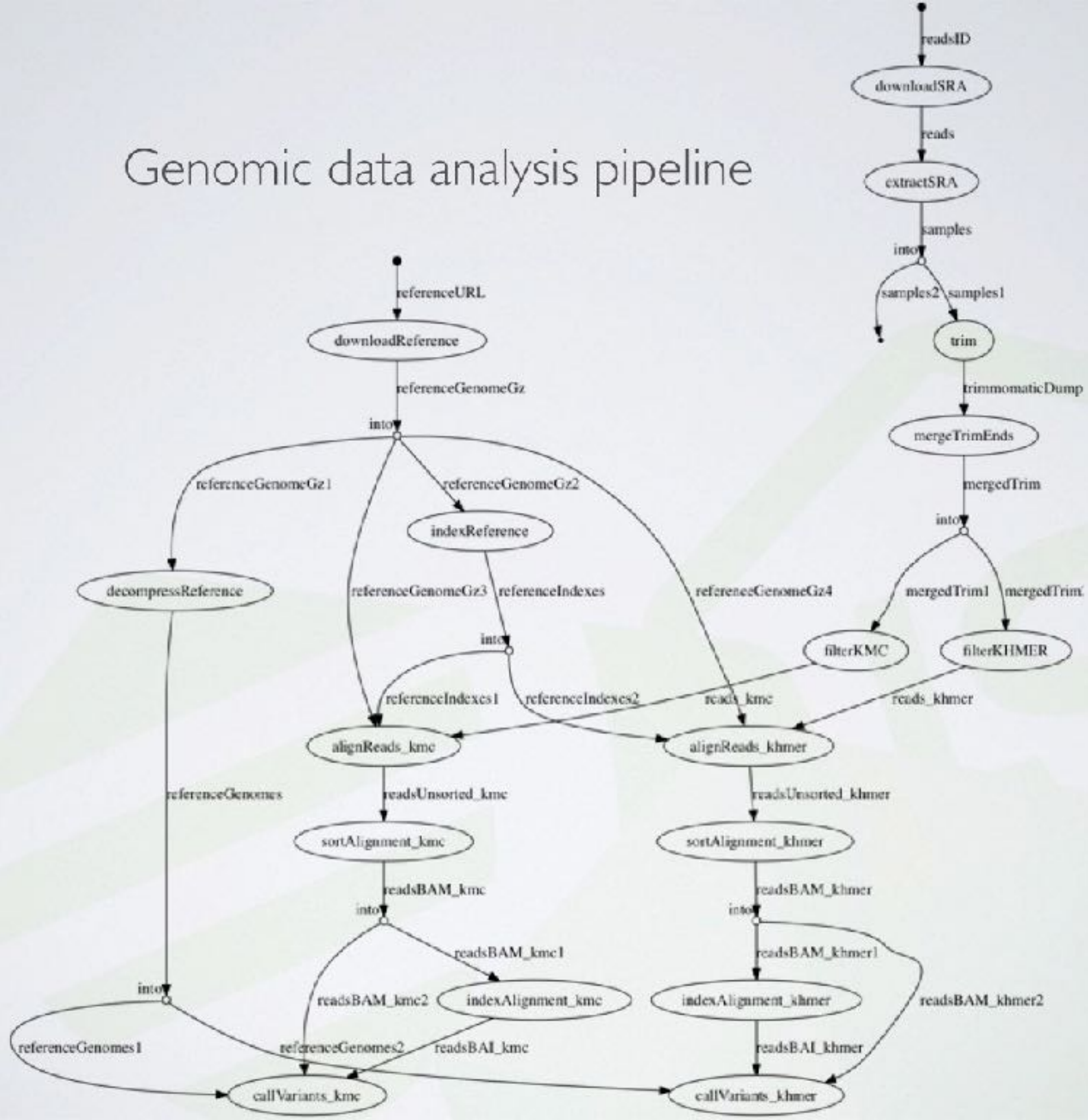




# PASSTHROUGH STREAM



# Genomic data analysis pipeline



# bionode-watermill

A Streaming Workflow Engine for bioinformatics and other big data explorations



Google Summer of Code 2016



# OIBIF

```
const samples = task({
  input: {
    db: 'sra',
    accession: config.sraAccession
  },
  output: '**/*.sra'
}, ({ input }) => ncbi.download(input.db, input.accession) )
```

```
const fastqDump = task({
  input: new File('**/*.sra'),
  output: [1, 2].map(n => new File(`*_${n}.fastq.gz`))
}, ({ input }) => shell(`fastq-dump --split-files --skip-technical --gzip ${input}`) )
```

# BIONODE COMMUNITY

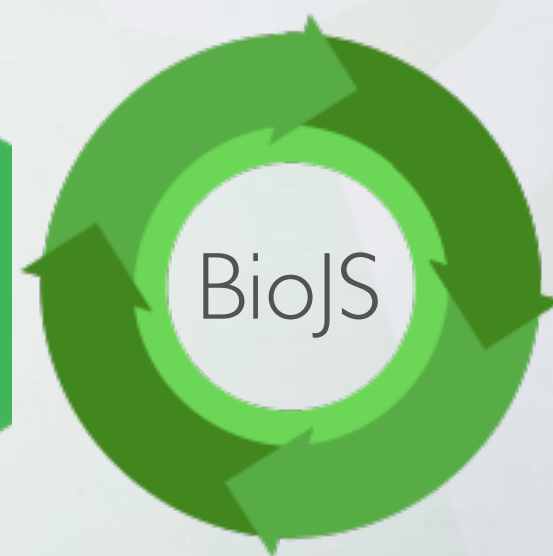


**WurmLab**  
**.github.io**



Queen Mary  
University of London

DNA  
digest







- **More self-sustainable**

- Organise more hackathons to attract contributors, especially with more technical Node/JS background
- Improve documentation and online tutorials to attract more users
- Apply for grants to hire a few full-time developers to improve and maintain the project

- **More useful**

- Reach out to specific labs/institutes to gather their needs and how this project can help them, and what features are missing
- Keep improving and working on existing code

- **More collaborative**

- Use our existing connections to other projects like Dat (data versioning and distribution), BioJS (biological visualisation), CWL (common workflow language) to see if together we can build something amazing for Open Genomics
- Use the Mozilla fellowship to make new collaborations



SUNDAY 11:00

**bionode.io**

**@bmpvieira**

## ACKNOWLEDGMENTS

**Community**

**Research group**

**Funding**



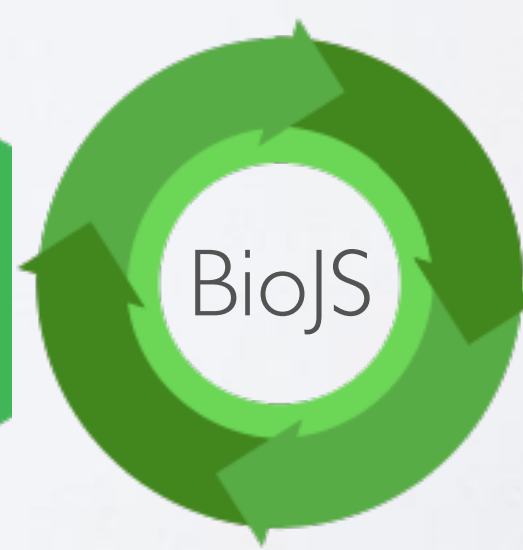
**WurmLab**  
**.github.io**



THE LEONA M. AND HARRY B.  
**HELMSLEY**  
CHARITABLE TRUST



**Friends**



# IMAGES SOURCES

<http://opengameart.org/content/bevouliin-green-flappy-bird-sprite-sheets>

<http://neoriceisgood.deviantart.com/art/100-furniture-sprites-405058884>

<http://android272.deviantart.com/art/Teeworlds-Teleport-570298308>

<http://www.how-to-draw-cartoons-online.com/eye-of-sauron.html>