

Dash Bio

Dash Components for Bioinformatics

Shammamah Hossain, software engineer
sham@plot.ly

10 Jul 2019



```
$ pip install dash-bio  
$ pip install dash-bio-utils
```

Overview

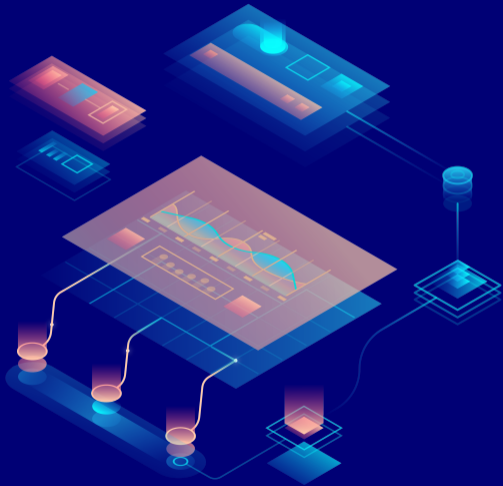
- Introductions to Dash and bioinformatics
- Motivations for the creation of `dash-bio`
- Dash Bio components
- `dash-bio-utils`: auxiliary library for compatibility with common file types
- Demonstrations

Introduction



What is Plotly?

- Plotly is a company
- ... But it's also the name of our open-source graphing libraries!
- Dash ("by Plotly") is a product by Plotly that we're going to talk about today



Dash

Dash lets you build beautiful web-based analytic interfaces on top of Python or R code.

- Interactive, web-based visualization
- Core libraries: dash, dash-core-components, dash-html-components
- Specialized libraries: dash-daq, dash-bio
- MIT licensed and open-source



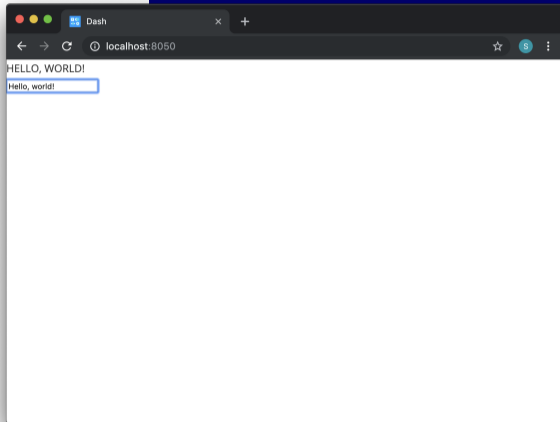
Creating a Dash App

```
import dash
import dash_html_components as html
import dash_core_components as dcc

app = dash.Dash()
app.layout = html.Div(children=[
    html.Div(id='output-div'),
    dcc.Input(id='text-input')
])

@app.callback(
    dash.dependencies.Output(
        'output-div', 'children'
    ),
    [dash.dependencies.Input(
        'text-input', 'value'
    )]
)
def capitalize_user_input(text):
    return text.upper()

app.run_server()
```

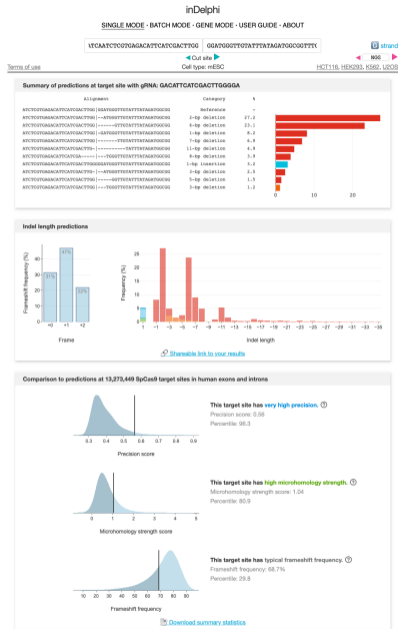


"I wanted to create a web app that is a joy to use. As a Python user, Dash was the obvious choice.

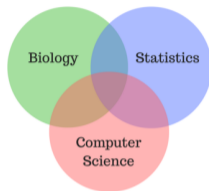
Not only has Dash been a great support for my scientific work, Dash has been artistically fulfilling to use.

It was amazing to build a functioning web app in just a handful of days, especially one that was slick and beautiful straight out of the box."

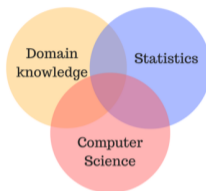
Max Shen, inDelphi



Bioinformatics



Data Science



Parallels between bioinformatics and data science.¹

Bioinformatics

- Interdisciplinary
- Typically employed in medical research
- e.g., genomics, proteomics, clinical trial analysis

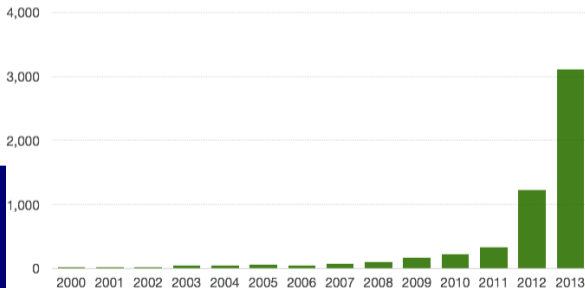
¹ <http://omgenomics.com/what-is-bioinformatics/>

Motivations



Why Bioinformatics?

- Rapid growth in the field
- Increasing use of “big data”

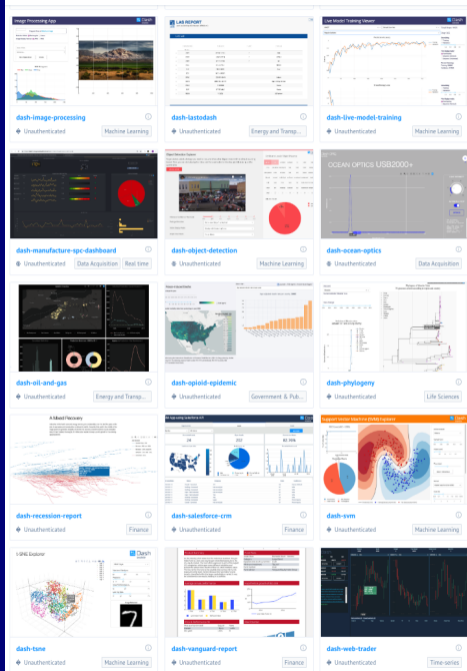


The number of bioinformatics papers by year from 2000 that mention big data.²

²<http://www.acgt.me/blog/2014/2/12/the-growth-of-bioinformatics-papers-that-mention-big-data>

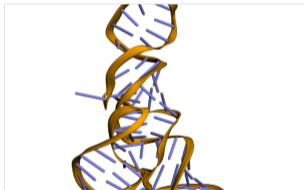
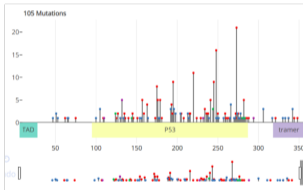
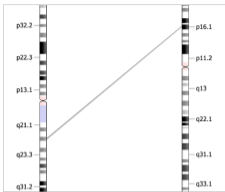
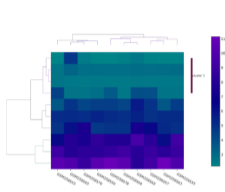
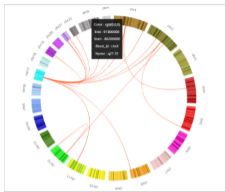
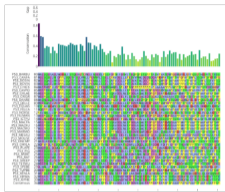
Why Dash?

- Bioinformaticians need highly specialized data analysis tools
- Dash components are “lego blocks” that can be used to create customized apps
- Libraries like GenomeDiagram, biopython, and GEOparse are already used by bioinformaticians



Dash Bio Components



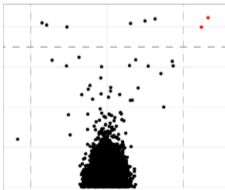
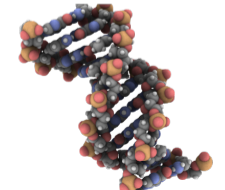
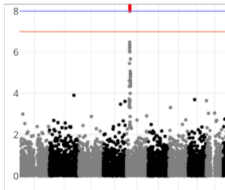
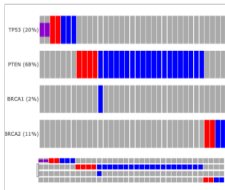


110 Insulin

Search in sequence.. (Regex supported)

```

1 MALWMRLLPL LALLALWGPD PAAAFVNOEL CGSHLVEALY
41 LVCGERGFYTPKTRREAE LQVGVDELGG GPGAGSLQPE
81 LEGHQRKG IYROCCSTIC SLVQLNRYC
  
```





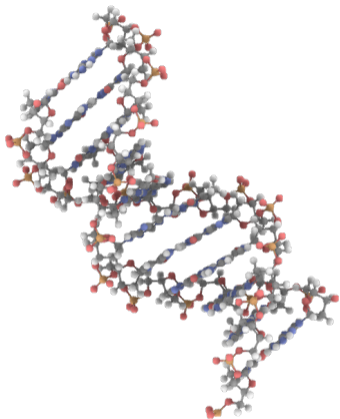
3D Rendering Tools



Custom Chart Types



Sequence Analysis
Tools



Speck

<https://github.com/wwwtyro/speck>

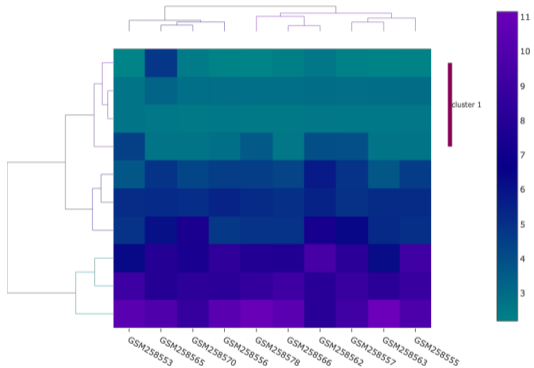
- WebGL-based molecule renderer
- Zoom/pan/rotate capabilities
- Ambient occlusion/outlines give a good sense of depth

Molecule3D

<https://github.com/autodesk/molecule-3d-for-react>

- Sticks, spheres, and cartoon representations of biomolecules
- Zoom/pan/rotate capabilities
- Selection by clicking: residue, atom, or chain





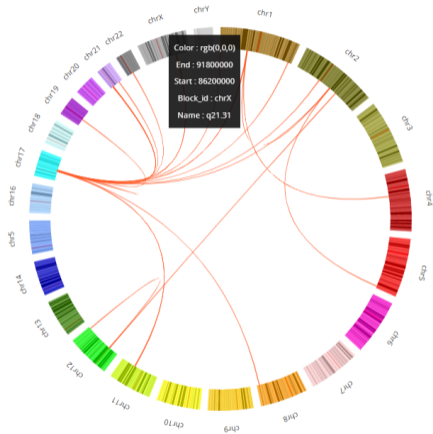
Clustergram

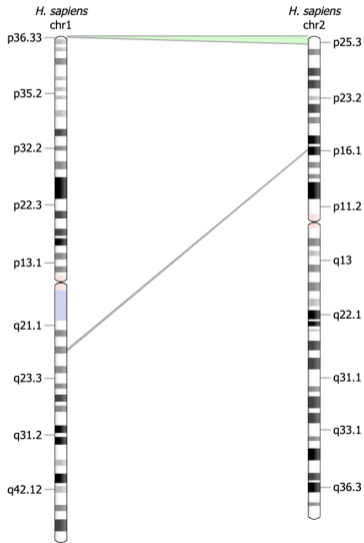
- Hierarchical clustering analysis within the component
- Traces grouped by cluster
- Configurable colors, clustering thresholds, and annotations

Circos

<https://github.com/nicgirault/circosJS>

- Circular representation of data
- Heatmaps, scatter plots, histograms
- Chords can connect different points on the circular layout





Ideogram

<https://github.com/eweitz/ideogram>

- Chromosome band visualization
- Homology feature
- Support for annotations and histograms

Sequence Viewer

<https://github.com/FlyBase/react-sequence-viewer>

- Subsequence search with regular expressions
- Clickable annotations with user-specified data
- Information about mouse selection

110 Insulin

Search in sequence.. (Regex supported)

```
1 MALWMRLPL LALLALWGPD PAAAFVNOHL CGSHLVEALY  
41 LVCGERGFY TPKTRREAED LQVGQVELGG GPGAGSLQPL  
81 ALEGSQKRG IVEQCCTSI SLYOLENYC
```

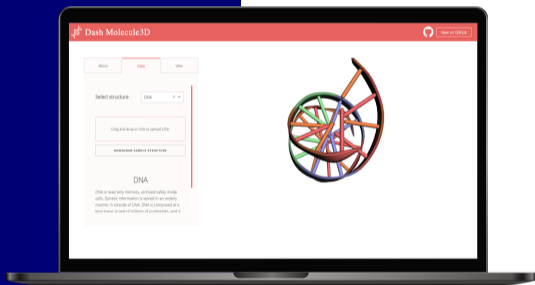
Parsing Tools



dash-bio-utils

- Parsers for data files
- Used for compatibility with dash-bio components
- Supports data files from popular databases
- SOFT, FASTA, PDB, XYZ, TSV

 nextprot UniProt GEO
Gene Expression Omnibus



Motivations

- Make development process faster
- Make Dash Bio app code more succinct
- Allow for a direct “database-to-visualization” workflow
- Leverage Dash App interface

<https://dash-gallery.plotly.host/dash-molecule3d/>

Demonstrations



Is Plotly free?

- Everything shown in this presentation is **free, open sourced**, and **MIT licensed**
- Our open-source development is funded by companies that request a specific feature or capability to add to Dash
- We also offer Dash Enterprise!

Dash Open Source

Analytic Application UI
Dash Open Source

Authorization & Authentication
AD & SAML Auth Middleware

Professional Grade Styling
Design Kit

PDF Reporting & Saved Views
Snapshot Engine

Advanced App Architectures
Redis Database & Python Task Queue

Continuous Deployment
Docker-based App Orchestrator

Data Monitoring
Scheduled Checks & PDF Reports

User Analytics
Built-in User Database & Usage Dashboard

Dash Enterprise

Resources

- Plotly website: <https://plot.ly>
- Dash Bio library: <https://github.com/plotly/dash-bio>
- Dash Bio demo apps:
<https://dash-gallery.plotly.host/Portal/?search=Bioinformatics>
- Dash Bio documentation: <https://dash.plot.ly/dash-bio>
- Dash documentation: <https://dash.plot.ly>
- Come visit us - we have a booth!

Questions?