The Geogenomic Mutational Atlas of Pathogens (GoMAP) Web System

David P. Sargeant, Michael W. Hedden, Sandeep Deverasetty, Christy L. Strong, Izua J. Alaniz, Alexandria N. Bartlett, Nicholas R. Brandon, Steven B. Brooks, Frederick A. Brown, Flaviona Bufi, Monika Chakarova, Roxanne P. David, Karlyn M. Dobritch, Horacio P. Guerra, Kelvy S. Levit, Kiran R. Mathew, Ray Matti, Dorothea Q. Maza, Sabyasachy Mistry, Nemanja Novakovic, Austin Pomerantz, Timothy F. Rafalski, Viraj Rathnayake, Noura Rezapour, Christian A. Ross, Steve G. Schooler, Sarah Songao, Sean L. Tuggle, Helen J. Wing, Sandy Yousif, Martin R. Schiller*

School of Life Sciences, University of Nevada Las Vegas, Las Vegas, Nevada, United States of America

Abstract

We present a new approach for pathogen surveillance we call Geogenomics. Geogenomics examines the geographic distribution of the genomes of pathogens, with a particular emphasis on those mutations that give rise to drug resistance. We engineered a new web system called Geogenomic Mutational Atlas of Pathogens (GoMAP) that enables investigation of the global distribution of individual drug resistance mutations. As a test case we examined mutations associated with HIV resistance to FDA-approved antiretroviral drugs. GoMAP-HIV makes use of existing public drug resistance and HIV protein sequence data to examine the distribution of 872 drug resistance mutations in ~502,000 sequences for many countries in the world. We also implemented a broadened classification scheme for HIV drug resistance mutations. Several patterns for geographic distributions of resistance mutations were identified by visual mining using this web tool. GoMAP-HIV is an open access web application available at http://www.bio-toolkit.com/GoMap/project/

Competing Interests: The authors have declared that no competing interests exist.

* E-mail: martin.schiller@unlv.edu

Introduction

There are a number of computation tools for visualizing pathogen surveillance data on world atlases. An example is the NCHHSTP Atlas web application at the United States Center for Disease Control and Prevention (CDC; www.cdc.gov). This, and similar applications are useful in managing responses to pathogen outbreaks. There are several examples of pathogens that have acquired mutations that make them resistant to one or more drugs. Indeed, this is a serious concern for treatment of patients infected with *Staphylococcus aureus*, *Clostridium difficile*, *Streptococcus pyogenes*, *Mycobacterium tuberculosis*, *Salmonella*, *Escherichia coli*, Influenza viruses, Human Immunodeficiency Virus (HIV), and several other viral infections. As sequencing costs plummet, pathogen genotyping is likely to become a routine diagnostic tool. Thus, examining the geographic distribution of mutations that are associated with drug resistance has the potential to contribute to the management of pathogen treatments. We call this approach Geogenomics.

To demonstrate the usefulness and feasibility of Geogenomics, we built a new computational system called Geogenomic Mutational Atlas of Pathogens (GoMAP). Since our first application is to HIV, the version presented herein is called GoMAP-HIV. We chose HIV for several reasons. Approximately 40 million people are infected with HIV worldwide, presenting a daunting health care challenge [1] for which there are over 25 drug compounds approved by the Food and Drug Administration (FDA) [2]. Although Highly Active Antiretroviral Therapy (HAART) provides an effective approach to treating HIV infection, there is significant *de novo* and transmission based drug resistance, especially if the treatment regimen is followed imperfectly [3–5]. For example, the rate of drug resistance in drug-naïve populations of Africa, East and Southeast Asia, North America, and Europe ranges from 5–12% [6].

Several databases (Stanford Database, World Health Organization (WHO), and Los Alamos) have consolidated many of the currently known HIV drug resistance mutations (DRMs). Another source of HIV data is the HIV genome sequences from many different patient blood samples housed in a number of databases. The largest of these is the Los Alamos HIV sequence database, which contains the genomic sequences from different patient blood samples encoding ~502,000 sequences of HIV proteins. While these two large datasets each provide a valuable resource, both can be used together to identify the occurrence of individual mutations in individual countries; however, this has not yet been reported.

The presence of HIV drug resistance mutations in the human population can be geographically assigned and the results may be useful for tracking the spread of DRMs, educating about HIV drug resistance, advancing our understanding of HIV evolution, and potentially for the management and design of therapies. To this end, we present the GoMAP-HIV open-access web applica-

tion, which provides a visual mining tool to help scientists better understand the global distribution of HIV DRMs.

Results

Data sources for the GoMAP-HIV database

Using the HIVToolbox MySQL database as a starting point, 1571 known HIV-1 DRMs (872 for FDA-approved drugs) were added from the Los Alamos and Stanford HIV databases, World Health Organization website, International AIDS Society and primary literature [7–9]. The HIVToolbox database has ~502,000 total sequences from patient blood samples in 126 different countries [10]. The appearance of DRMs in a country was determined by performing multiple sequence alignments of protein sequences from patient blood samples with an HIV reference protein as previously created using ClustalΩ to generate clade specific alignments in the HIVToolbox database [11].

GoMAP-HIV User Interface

The main interface is composed of two primary sections, the Naïve and Treated Atlas Window and the DRM Selection and Display Tools.

Naïve and Treated Atlas Window. This window consists of two panels, each displaying data for naïve or treated patients on an atlas of all recognized countries (**Fig. 1**). When a user first enters the GoMAP-HIV site, the countries in both atlases are initially white. Hovering the mouse over each country results in the appearance of a label containing the country name (**Fig. 1A**). The user can zoom in or out of each of the atlases using the plus and minus icons located in the top left corner of each panel.

The user can first select a country first or a drug. Upon selection of a country, antiretroviral (ARV) drug, and then a DRM from the selection windows, color-coded countries are displayed on the atlases (**Fig. 1**). The colors are: white indicates no observed DRM, red indicates at least one sequence containing the DRM, and gray indicates those countries that do not have sequences in our database. The left atlas panel is driven by patient blood sample sequence data obtained from treatment-naïve patients while treatment-experienced patient blood sample data drives the right panel (**Fig. 1A and 1B**, respectively). In order to populate the atlases with the DRM data, one must first choose from the DRM

Selection Tools found below the atlas panels. The following are found under the DRM Selection Tools:

Country Selection window. This selector consists of seven expandable categories organized by geographic region. When the user first enters the GoMAP-HIV site, all countries are selected as indicated by the checkbox in the tile bar (**Fig. 2A**). The "All Countries (126 with data)" category consists of an alphabetical list of the countries in the database that have associated DRMs. Thus far, we have HIV sequence data from 126 out of the 253 recognized countries of the world. The other six categories contain the same country information, but the countries are grouped by continent. To access DRM data for a specific country, the user selects that country by clicking on the country name found in either the "All Countries" or continent category, or by using the mouse to click on a country in the atlas displays. Once the country selection has been made, this updates information into the Drug Selection window. At any time, checking the "Select All Countries" checkbox will update data for all countries in the world.

Drug Selection window. Initially, this window contains a list of 23 FDA-approved drugs with the total number of sequences for the target of each drug. This window is updated after selection of a country. The title bar displays the name of the selected country. The Drug Selection window is populated with up to 23 FDA approved drugs that have sequence data for the HIV protein drug target (**Fig. 2B**). The FDA drugs are color-coded based on inhibitor type, with the color-code legend located at the bottom of the drug selection window. The first column of the table in this window contains the drug names. The second and third columns contain the number of naïve-patient sequences or treated-patient sequences for the country selected. Selection of a drug from the first column triggers two events: 1) indication of the drug selected in the DRM selection Title bar, and 2) population of the DRM Selection window with DRMs for the ARV selected.

DRM Selection window. Each row of the DRM selection table shows information about a DRM for the country and drug previously selected (**Fig. 2C**). Once a DRM is selected with a mouse click, three events are triggered: the DRM Selection title bar shows the selected DRM and drug, the atlases are color coded for the presence of the DRM in each country, and the DRM Observation by Country table is populated (**Fig. 2D**).
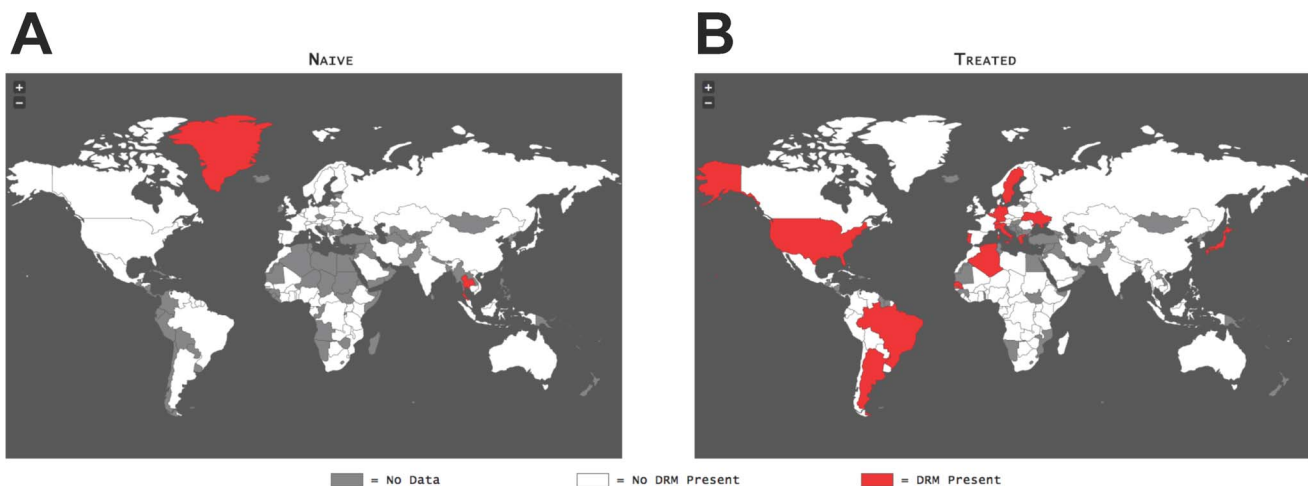

Figure 1. GoMAP example maps. World atlas displays showing observation of DRMs derived from patient blood samples for naïve (**A**) and treated (**B**) patients. The examples shown are for the beneficial N88S DRM for Fosamprenavir. A color legend for country coloring is shown (bottom).
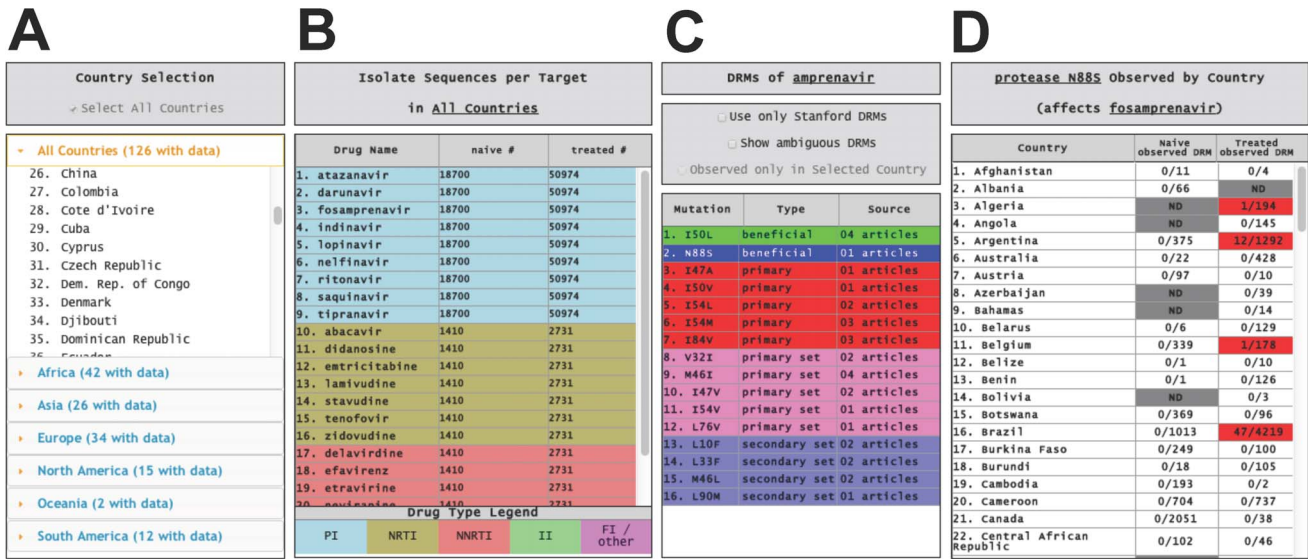doi:10.1371/journal.pone.0092877.g001

A. Country Selection

☑ Select All Countries

All Countries (126 with data)
- 26. China
- 27. Colombia
- 28. Cote d'Ivoire
- 29. Cuba
- 30. Cyprus
- 31. Czech Republic
- 32. Dem. Rep. of Congo
- 33. Denmark
- 34. Djibouti
- 35. Dominican Republic
- 26. Ecuador

▸ Africa (42 with data)
▸ Asia (26 with data)
▸ Europe (34 with data)
▸ North America (15 with data)
▸ Oceania (2 with data)
▸ South America (12 with data)

B. Isolate Sequences per Target in All Countries

Drug Name	naive #	treated #
1. atazanavir	18700	50974
2. darunavir	18700	50974
3. fosamprenavir	18700	50974
4. indinavir	18700	50974
5. lopinavir	18700	50974
6. nelfinavir	18700	50974
7. ritonavir	18700	50974
8. saquinavir	18700	50974
9. tipranavir	18700	50974
10. abacavir	1410	2731
11. didanosine	1410	2731
12. emtricitabine	1410	2731
13. lamivudine	1410	2731
14. stavudine	1410	2731
15. tenofovir	1410	2731
16. zidovudine	1410	2731
17. delavirdine	1410	2731
18. efavirenz	1410	2731
19. etravirine	1410	2731
20. nevirapine	1410	2731

Drug Type Legend: PI | NRTI | NNRTI | II | FI / other

C. DRMs of amprenavir

☐ Use only Stanford DRMs
☐ Show ambiguous DRMs
☐ Observed only in Selected Country

Mutation	Type	Source
1. I50L	beneficial	04 articles
2. N88S	beneficial	01 articles
3. I47A	primary	01 articles
4. I50V	primary	01 articles
5. I54L	primary	02 articles
6. I54M	primary	03 articles
7. I84V	primary	03 articles
8. V32I	primary set	02 articles
9. M46I	primary set	04 articles
10. I47V	primary set	02 articles
11. I54V	primary set	01 articles
12. L76V	primary set	01 articles
13. L10F	secondary set	02 articles
14. L33F	secondary set	02 articles
15. M46L	secondary set	02 articles
16. L90M	secondary set	01 articles

D. protease N88S Observed by Country (affects fosamprenavir)

Country	Naive observed DRM	Treated observed DRM
1. Afghanistan	0/11	0/4
2. Albania	0/66	ND
3. Algeria	ND	1/194
4. Angola	ND	0/145
5. Argentina	0/375	12/1292
6. Australia	0/22	0/428
7. Austria	0/97	0/10
8. Azerbaijan	ND	0/39
9. Bahamas	ND	0/14
10. Belarus	0/6	0/129
11. Belgium	0/339	1/178
12. Belize	0/1	0/10
13. Benin	0/1	0/126
14. Bolivia	ND	0/3
15. Botswana	0/369	0/96
16. Brazil	0/1013	47/4219
17. Burkina Faso	0/249	0/100
18. Burundi	0/18	0/105
19. Cambodia	0/193	0/2
20. Cameroon	0/704	0/737
21. Canada	0/2051	0/38
22. Central African Republic	0/102	0/46

Figure 2. GoMAP country, drug, and DRM data selectors. A. The Country Selection window is used to load data into world atlas displays and populate the Drug Selection table. **B.** The Drug Selection window is for choosing an FDA-approved drug. ARVs are grouped by type: Protease inhibitors (PI) are colored a light blue, nucleoside/nucleotide reverse transcriptase inhibitors (NRTI) are tan, non-nucleoside reverse transcriptase inhibitors (NNRTI) are coral, integrase inhibitors (II) are light green, and fusion inhibitors (FI) are light magenta. The number of different patient blood samples for a country selected in A is shown. If the "Select All Countries" checkbox is checked, data for all countries are shown. Selection of a drug populates the DRM Selection table with DRM data for that drug, and country if selected. **C.** The DRM Selection table shows DRMs ordered by their classification. Once a DRM is selected (N88S here) the presence of DRMs in each country are plotted onto the world atlases and the DRM Observation By Country table (**D**) is populated. The DRM Selection table also has a hyperlink to one or more published papers supporting the DRM. Checkboxes below the title bar enable display of DRMs present in other databases for which no published scientific paper could be identified (designated ambiguous), for showing the Major/Minor DRM classification scheme, and for examining DRMs only in the country selected. **D.** The DRM Observation By Country table shows a list of the observed DRMs (numerator) and total sequences examined (denominator) for each country. The data are stratified for naïve and treated patients.
doi:10.1371/journal.pone.0092877.g002

The DRM Selection table contains three columns with information about the DRM, type of DRM, and links to PubMed abstract(s) containing experimental evidence that supports the DRM and its classification (**Fig. 2C**). The first column in the table contains a color-coded list of the DRMs related to the selected drug. The IUPAC signal letter code for the wild-type amino acid, the amino acid position, and the identity of the mutated amino acid are shown (e.g. I47A) [12].

The type of DRM is displayed in the second column and each row is colored using a DRM classification scheme. GoMAP-HIV contains two different DRM classification schemes that can be chosen from a checkbox just below the DRM selection title bar. A subset of the data in the GoMAP-HIV database comes from the Stanford HIV Drug Resistance Database and had DRM classifications of major or minor [13]. GoMAP-HIV contains 188 DRMs from the Stanford Database that have one of these designations and selection of the checkbox below the title bar restricts the DRM selections to these DRMs.

As the default classification scheme, we provide a broader approach to classification. We implemented this new scheme because as we annotated DRMs from the literature and other databases, we observed DRMs that did not fit into the standard categories of major and minor [13]. Definitions for the new scheme can be found in **Table 1**. Briefly, DRM types designated *beneficial* or *beneficial set* (decrease drug resistance) are colored different shades of green. Those that cause resistance, *primary* and *primary set*, are colored red and pink, respectively. Those that amplify resistance are called *secondary set* and are colored purple. The few mutations that do not affect resistance directly, but which are precursors to other DRMs are called *precursors* and colored light blue. There is a checkbox option to view *ambiguous* mutations, which are colored white. Ambiguous mutations are those DRMs identified from another database where a published peer-reviewed primary source could not be identified.

The DRM list is initially populated with all DRMs observed for the drug selected. However, using the mouse to click the "Observed only in selected Country" checkbox in the title bar will repopulate the DRM list showing only those DRMs observed in the country selected.

DRM Observation by Country window. Once a drug has been selected from the DRM Selection window, the title bar changes to include the DRM and drug (**Fig. 2D**). The first of three columns in this table contains an alphabetized list of all countries with DRMs related to the selected drug. Each row in the second and third columns indicates whether the DRM has been observed in naïve or treated patient blood samples from that country, respectively. In each column the numerator is the number of sequences in the country that have the DRM selected and the denominator is the total number of sequences in the country indicated. The information for the naïve- or treated-DRM is color-coded to match the atlas displays.

Example GoMAP global analysis of Fosamprenavir DRMs

GoMAP is the first tool that can be used to look at the distribution of individual HIV DRMs for HIV drugs. **Figs. 3** and **4** shows examples of geographic distribution plots generated for several DRMs in HIV protease previously associated with resistance to Fosamprenavir. For protease, there are 90,474 sequences from 120 countries. There are 16 different mutations in HIV protease that influence resistance to Fosamprenavir includ-

Table 1. Definitions for different types of drug resistance mutations.

Type of DRM	Definition
Primary	Causes resistance without any other mutations
Primary set	Two or more mutations that cause resistance only in the presence of other primary set mutation(s)
Secondary	Enhances resistance caused by a primary mutation
Resistance precursor	A mutation that has no effect on resistance, but must occur prior to another primary or primary set of mutations
Beneficial	A mutation that prevents or reduces resistance
Beneficial set	Two or more mutations that when occurring simultaneously prevent or reduce resistance
Ambiguous	Cannot determine the nature based on existing literature

ing: primary mutations I47A, I50V, I54L, I54M, I84V; primary set mutations V32I, M46I, I47V, I54V, and L76V, beneficial mutations I50L and N88S; and secondary set mutations L10F, L33F, M46L, and L90M [14–26].

The global distributions of individual DRMs for Fosamprenavir in naïve patients show interesting patterns (**Fig. 3**). As expected the 5 primary mutations were not ubiquitously observed, each showing up in only 0–3 countries. The two beneficial mutations had similar profiles. Of the 5 primary set mutations, 2 were not yet observed (V32I and I47V), but the other 3 were prevalent showing up in countries from most continents. In particular L76V was the most prevalent of the DRMs, observed in most countries of Europe, the Americas, and Australia.

In general, the global distribution of Fosamprenavir DRMs for patients previously receiving ARV therapy showed more wide spread prevalence (**Fig. 4**). Primary DRM I84V was the most widespread with occurrences in most countries in Europe, the

Americas, and Australia. I47A has only been observed in Venezuela and other primary DRMs were observed in several countries in many continents. One major noticeable difference was for primary set mutations, with all showing generally widespread distributions on most continents.

Discussion

The CDC, WHO and other agencies involved in managing the impact of pathogens on human health all have programs that focus on surveillance. These data, when mapped, assist in the investigation and management of outbreak events and epidemics. Indeed, surveillance studies have led to the identification of a growing concern in pathogen management. Since the landmark work documenting Methicillin-resistant *Staphylococcus aureas* (MRSA) in hospitals of Great Britain in the 1960s it has become clear that strains of pathogens can become resistant to antibiotics
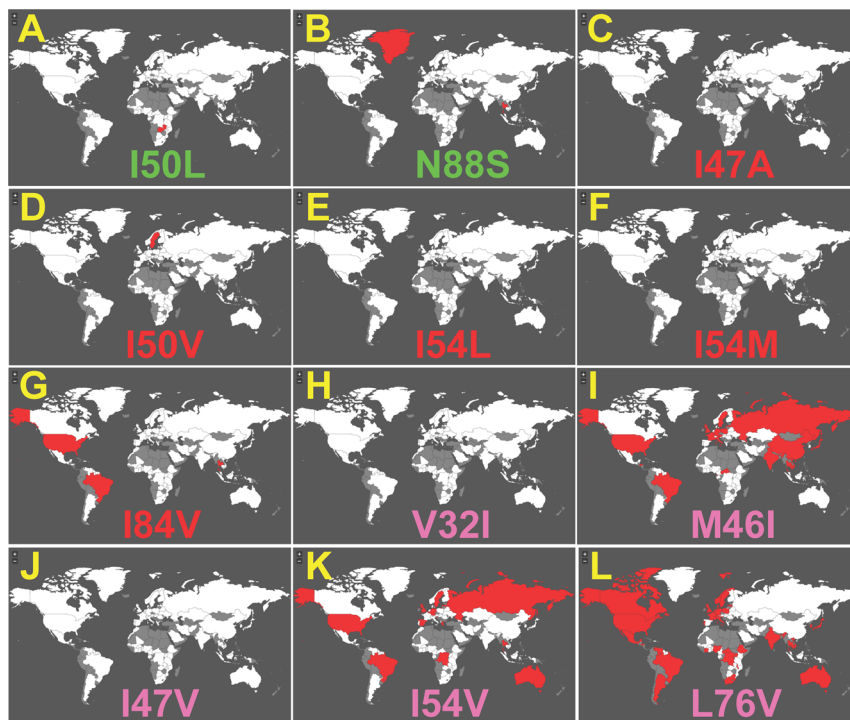

Figure 3. GoMAP example maps showing occurrence of DRMs for Fosamprenavir in genomes of naïve patients. Atlas displays for all 12 beneficial, primary, and primary set DRMs in the protease gene associated with Fosamprenavir and only for patient blood samples from drug-experienced patients (**A-L**). DRM type labels are colored as in the DRM table (**Fig. 2**).
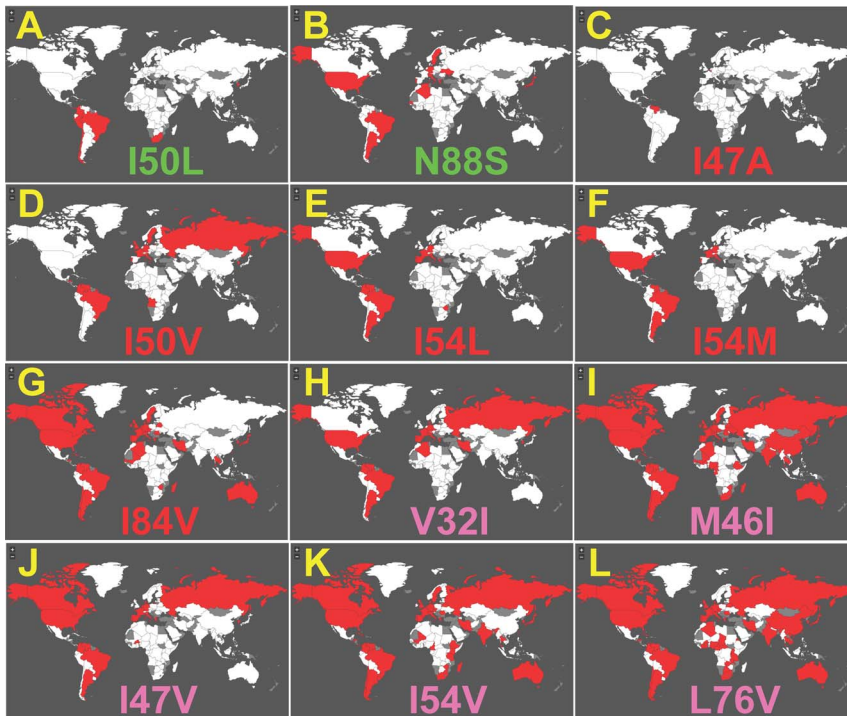doi:10.1371/journal.pone.0092877.g003

Figure 4. GoMAP example maps showing occurrence of DRMs for Fosamprenavir in genomes of drug experienced patients. Atlas displays for all 12 beneficial, primary, and primary set DRMs in the protease gene associated with Fosamprenavir and only for patient blood samples sequences from naïve patients (**A-L**). DRM type labels are colored as in **Fig. 3**.
doi:10.1371/journal.pone.0092877.g004

[27]. This has become more predominant with a number of bacterial and viral pathogens acquiring DRMs. Although there is a growing body of DRM data for a variety of pathogens, a tool that relates this data to geographic surveillance is missing. By combining the examination of geographic distribution of pathogen cases with the appearance of drug resistance mutations, we propose a new scheme that should aid in the management of pathogen-borne illness. Because this method relies on detecting drug resistance mutations in the genomes of pathogens and metadata for the geographic location of the sample, we call this new approach "Geogenomics".

In order to demonstrate the feasibility and utility of Geogenomics, we sought to first examine a pathogen that had extensive data on genome sequences with geographical locations. HIV-1 met these criteria and we were able to obtain and use sequence-based data from the public Los Alamos HIV sequence database. We have built a Java-based web application called GoMAP. The GoMAP-HIV implementation presented herein focuses on HIV.

In addition to the sequence data, we needed a source of HIV-1 DRMs. We first attempted to use data sources from existing DRM databases, but noted that some DRMs reported in these databases were not supported by a published peer-reviewed science paper, a concept we have enforced in previous database/web systems we have built [28–30]. We therefore generated a consolidated HIV DRM database with records supported by at least one published paper (see methods). This annotation effort resulted in 1571 DRMs, 872 of which are for 23 known FDA-approved ARVs for treating HIV-1 infection.

23 different HIV ARVs can be investigated with GoMAP (Maraviroc is not included because GP120 is too polymorphic, nor is Dolutegravir due to its recent release to the market). We present a test case analysis of Fosamprenavir as an example of GoMAP-HIV (**Figs 3.** and **4**). We consistently observed the presence of

primary DRMs in the drug-experienced patients from a small set of countries mostly clustered in central Europe. This geographic distribution is consistent with several studies previously reported. One study found that 29% of the French patients that were enrolled in the study and receiving Amprenavir/Ritonavir were drug resistant [31]. A second study found that 37% of the German patients enrolled in the study, who were previously treated with another protease inhibitor, were also resistant to Amprenavir [32]. Note that Fosamprenavir is a prodrug of Amprenavir and metabolized into Amprenavir in the intestine [33]. The example analysis of Fosamprenavir demonstrates the potential value in identifying countries where HIV sequences have DRMs that are known to be associated with resistance to specific drugs.

For GoMAP-HIV, we wanted to be as comprehensive as possible in our presentation of HIV DRMs. On the web system, we provide the standard major/minor classification scheme used by the IAS and Stanford HIV Drug Resistance database. However, we identified some limitations in this scheme. The designation "major" is defined as "mutations that reduce drug susceptibility by themselves", and the term minor is used for "mutations that reduce drug susceptibility or improve the replicative fitness of isolates with a major mutation"[13]. In GoMAP, we have added additional DRM definitions to be inclusive of other types of mutations that we observed as we annotated the DRM data and set criteria for such DRM designations.

Major, occasionally called primary, is similar to what we are now calling *primary*. For primary DRMs in the GoMAP annotations, the mutation must be observed in a clinical patient blood sample that demonstrates significant drug resistance and when introduced into a clean viral background, produces significant resistance in culture without the presence of other mutations. We recognized that some DRMs only produce

resistance when in combination, which does not fit well into the major and minor scheme. We call these *primary sets*.

Minor, occasionally called secondary, is similar to what we are now calling *secondary*. We also recognized the need for *secondary sets*, where sets of mutations are identified that enhance viral fitness when combined with a primary mutation. We also needed classifications that did not previously exist for *beneficial, beneficial set*, and *precursor* for mutations that did not fit into the major/minor classification scheme. This scheme is the default scheme used in GoMAP-HIV, but there is an option to use the Major/Minor scheme.

While GoMAP provides a system for examining a global view of HIV DRMs, we recognize that there are a number of limitations concerning the HIV data. One of the challenges with epidemiological studies is that there is a number of factors of concern in compiling data from different studies [34]. First, the sampling is biased. The different patient blood samples have been complied over a ~30 year period. Each individual study has variable patient selection, assay methods employed, and interpretation by the investigators. Second, when an isolate is sequenced, we must consider that it represents one sequence of many circulating forms of the virus that contain mutations. Third, some countries, like Cambodia, have a low number of sequenced patient blood samples. Nevertheless, our implementation is intended to provide an example of a new approach that can be used for tracking and managing pathogens.

Another limitation is that some HIV DRMs were only observed upon treatment with two or more drugs at once and are not used in our plots because in this situation the ambiguity in the drug causing the DRM cannot be resolved. Another concern is that certain clades are known to have relatively specific geographical distributions and DRMs shown on GoMAP may somewhat represent clade distributions. Although most HIV-1 clades are thought to have similar sensitivity to drug resistance, there are specific cases where certain clades are more prone to drug resistance, e.g. [6,35]. However, very few of our DRM maps show distributions that match previously published clade distributions [6].

We used GoMAP to mine for geogenomic DRM patterns. Although the data has limitations, it represents a rather large compendium of many studies with numerous sequences, and can provide some insight into the evolution of HIV and the geographic spread of particular DRMs.

We observed several patterns of regional concentration in DRMs that generally were not coincidental with the established global subtype distributions [36]. The I50L primary mutation in protease for Atazanavir and Amprenavir is predominant in South America, but not in the same countries as the B or BF subtype distributions. Likewise, the Y188H primary mutation in reverse transcriptase for Delavirdine and Nevirapine is in a cluster of 3 countries in central Africa in sequences for drug-experienced patients. The G190V primary mutation in reverse transcriptase for Nevirapine is in a small cluster of European countries. One obscure observation is that all 8 primary DRMs for Enfuvirtide are observed in Italy, and some of these primary DRMs are also in a few other counties; samples with viruses harboring some of these DRMs have been observed in Italian patients [37]. Collectively, the geographic patterns suggest a transmission-based spread for these DRMs.

The distribution of primary DRMs is of concern, because it could negatively impact ARV therapy, especially if the DRM is in naïve patients. Several primary DRMs had widespread global distributions. The I93L primary DRMs for Indinavir and L76V for Lopinavir are widespread, present in most countries, including naïve patients. Other primary DRMs with a global distribution are N89I, L90M for protease and E399D, V179I, and G196E for reverse transcriptase.

As expected, some DRMs are rarely observed in naïve patients, showing in only 0 or 1 country, but are widespread in patients that have undergone ARV therapy. These include V82T, V82S, L24I, and I47V for protease, and E44D, V106A, Y181V, H221Y and F227L for reverse transcriptase. This is typically observed in epidemiological studies, e.g. [38].

Surprisingly, a large number of published primary DRMs are not observed in any country for naïve or treated patients. These include DRMs (I47A, I84C) for protease, (P143S, K65N, K70E, K70G, M184T, Y181V, G190C, G190Q, F227C, K101P, K103T, K103H, I132M, Y181V, Y 188C, G190C, G190Q, E138R) for reverse transcriptase, and F121Y, Y143R, Y143C, and N155S for integrase. These mutations were observed in controlled experimentation; so why they have not appeared in this larger cohort of patients is unknown. One possibility is that these DRMs may provide resistance to ARVs in culture, but not so *in vivo*, perhaps due to differential recognition by the immune system.

The distribution of beneficial DRMs is of interest, because it could positively impact the efficacy of ARV therapy. We observed three beneficial mutations for Darunavir (K20I, E35D and V82A) that have a widespread global distribution, even in naïve patients. Another beneficial DRM with widespread distributions include A71V for Saquinavir. Notably the K70E beneficial DRM for Zidovudine was not observed in any country. However, the distribution of beneficial DRMs should be interpreted with caution because there are examples where a DRM is beneficial for one ARV, but a primary DRM for another ARV (e.g. N88S and L76V for protease).

The use of biogeography in tracking pathogens has recently emerged in several other laboratories with applications that are synergistic with GoMAP. Like the WHO and CDC, the Center for Disease, Dynamics, Economics and Policy also has a resistance map web system for bacterial pathogens that is synergistic with GoMAP (http://www.cddep.org/).

Another web tool, Healthmap advances surveillance mapping systems by integrating agency reporting, allowing rapid tracking of H1N1 Influenza outbreaks [39]. We considered enabling annual temporal tracking of HIV in GoMAP, since the Los Alamos HIV sequence database had meta data associated with the year that the sample was collected: unfortunately, the data were too sparse to generate useful maps. However, GoMAP can be adapted to include this type of feature. In a step toward mapping pathogen mutations, Supramap uses a similar concept for rapid tracking of pathogen spread but also includes genomics to identify and track movement of specific pathogen strains [40]. This can be used to examine how pathogens mutate over time and space. GoMAP is more specifically and synergistically focused on tracking the spread of drug resistance mutations for pathogens. Our implementation examining HIV provides a system and approach to examine HIV drug resistance mutations.

In conclusion, GoMAP embodies a new "Geogenomic" approach for examining pathogen mutations. We think this tool will be useful for examining the evolution of pathogens, examining how drug resistance spreads, and for those working in the area of public health. We offer our tool to public health agencies for their own use in tracking HIV mutations or mutations in other pathogens. GoMAP is an open access web application available at http://gomap.bio-toolkit.com/GoMap/. Documentation of Go-MAP and a User Guide are at http://www.bio-toolkit.com/GoMAP/project. The source code and SQL database of DRMs for HIV is available upon request.

Methods

GoMAP-HIV implementation

GoMAP-HIV was implemented as a Java-based web application using standard three-tier software architecture consisting of a relational database for information management layer, a Java Servlet-based backend for data access, and a web-based presentation layer that is built using AJAX technology. The data was created by aligning amino acid sequences from patient blood samples for each country using ClustalΩ (http://www.clustal.org) and then analyzing the aligned sequences for individual DRMs. The presentation layer includes browser-based Geographical Information System (GIS) libraries (e.g. JVectorMap, http://www.jvectormap.com) and uses bioinformatics libraries to read and work with alignment files and the resulting aligned sequences (e.g. Strap, http://www.bioinformatics.org/strap/; BioJava,

http://www.biojava.org). The JavaScript library jQuery (http://www.jquery.com) was also used in the presentation layer.

Author Contributions

Conceived and designed the experiments: MRS DPS CLS MWH. Performed the experiments: DPS MWH SD SGS CAR. Analyzed the data: IJA ANB NRB SBB FAB FB MC RPD KMD HPG KSL KM RM DQM SM NN AP TR VR NR SS SLT SY SGS MRS CLS SD. Wrote the paper: MRS DS CLS. Helped generate the video tutorials: HJW MRS.

References

1. Tanne JH (2006) Nearly 40 million people worldwide are infected with HIV. BMJ 332: 1289–1289. doi:10.1136/bmj.332.7553.1289.
2. Jain R, Clark NM, Diaz-Linares M, Grim SA (2006) Limitations of current antiretroviral agents and opportunities for development. Curr Pharm Des 12: 1065–1074.
3. Blaise P, Clevenbergh P, Vaira D, Moutschen M, Dellamonica P (2002) HIV resistance to antiretroviral drugs: mechanisms, genotypic and phenotypic resistance testing in clinical practice. Acta Clin Belg 57: 191–201.
4. Shekelle P, Maglione M, Geotz MB, Wagner G, Wang Z, et al. (2007) Antiretroviral (ARV) drug resistance in the developing world. Evid Rep Technol Assess (Full Rep): 1–74.
5. Geretti AM (2007) Epidemiology of antiretroviral drug resistance in drug-naïve persons. Curr Opin Infect Dis 20: 22–32.
6. Spira S, Wainberg MA, Loemba H, Turner D, Brenner BG (2003) Impact of clade diversity on HIV-1 virulence, antiretroviral drug sensitivity and drug resistance. J Antimicrob Chemother 51: 229–240.
7. Kagan R, Winters M, Merigan T, Heseltine P (2004) HIV type 1 genotypic resistance in a clinical database correlates with antiretroviral utilization. AIDS Res Hum Retroviruses 20: 1–9. doi:10.1089/088922204322749440.
8. Leitner T, Foley B, Hahn B, Marx P, McCutchan F, et al. (2005) HIV Sequence Compendium, 2005 Los Alamos, NM: Theoretical Biology and Biophysics Group, Los Alamos National Laboratory.
9. Johnson VA, Brun-Vézinet F, Clotet B, Günthard HF, Kuritzkes DR, et al. (2010) Update of the drug resistance mutations in HIV-1: December 2010. Top HIV Med 18: 156–163.
10. Sargeant D, Deverasatty S, Luo Y, Baleta AV, Zobrist S, et al. (2011) HIVToolbox, an integrated web application for investigating HIV. PloS One 6: e20122.
11. Sievers F, Wilm A, Dineen D, Gibson TJ, Karplus K, et al. (2011) Fast, scalable generation of high-quality protein multiple sequence alignments using Clustal Omega. Molecular Systems Biology 7. Available: http://www.nature.com/doifinder/10.1038/msb.2011.75. Accessed 20 June 2012.
12. IUPAC-IUB Commission on Biochemical Nomenclature (CBN) (1970) Abbreviations and symbols for nucleic acids, polynucleotides and their constituents. Recommendations 1970. BiochemJ 120: 449–454.
13. Shafer R (2002) Genotypic Testing for Human Immunodeficiency Virus Type 1 Drug Resistance. Clinical Microbiology Reviews 15: 247–277. doi:10.1128/CMR.15.2.247-277.2002.
14. Partaledis JA, Yamaguchi K, Tisdale M, Blair EE, Falcione C, et al. (1995) In vitro selection and characterization of human immunodeficiency virus type 1 (HIV-1) isolates with reduced sensitivity to hydroxyethylamino sulfonamide inhibitors of HIV-1 aspartyl protease. J Virol 69: 5228–5235.
15. Prado JG, Wrin T, Beauchaine J, Ruiz L, Petropoulos CJ, et al. (2002) Amprenavir-resistant HIV-1 exhibits lopinavir cross-resistance and reduced replication capacity. AIDS 16: 1009–1017.
16. Maguire M, Shortino D, Klein A, Harris W, Manohitharajah V, et al. (2002) Emergence of resistance to protease inhibitor amprenavir in human immunodeficiency virus type 1-infected patients: selection of four alternative viral protease genotypes and influence of viral susceptibility to coadministered reverse transcriptase nucleoside inhibitors. Antimicrob Agents Chemother 46: 731–738.
17. Sista P, Wasikowski B, Lecocq P, Pattery T, Bacheler L (2008) The HIV-1 protease resistance mutation I50L is associated with resistance to atazanavir and susceptibility to other protease inhibitors in multiple mutational contexts. J Clin Virol 42: 405–408. doi:10.1016/j.jcv.2008.03.023.
18. Yanchunas J Jr, Langley DR, Tao L, Rose RE, Friborg J, et al. (2005) Molecular basis for increased susceptibility of isolates with atazanavir resistance-conferring substitution I50L to other protease inhibitors. Antimicrob Agents Chemother 49: 3825–3832. doi:10.1128/AAC.49.9.3825-3832.2005.
19. Gatanaga H, Suzuki Y, Tsang H, Yoshimura K, Kavlick MF, et al. (2002) Amino acid substitutions in Gag protein at non-cleavage sites are indispensable

for the development of a high multitude of HIV-1 resistance against protease inhibitors. J Biol Chem 277: 5952–5961. doi:10.1074/jbc.M108005200.
20. Colonno R, Rose R, McLaren C, Thiry A, Parkin N, et al. (2004) Identification of I50L as the signature atazanavir (ATV)-resistance mutation in treatment-naive HIV-1-infected patients receiving ATV-containing regimens. J Infect Dis 189: 1802–1810. doi:10.1086/386291.
21. Cunyat F, Ruiz L, Marfil S, Puig T, Bofill M, et al. (2010) Genotypic and phenotypic evolution of HIV type-1 protease during in vitro sequential or concomitant combination of atazanavir and amprenavir. Antivir Ther (Lond) 15: 431–436. doi:10.3851/IMP1543.
22. Young TP, Parkin NT, Stawiski E, Pilot-Matias T, Trinh R, et al. (2010) Prevalence, mutation patterns, and effects on protease inhibitor susceptibility of the L76V mutation in HIV-1 protease. Antimicrob Agents Chemother 54: 4903–4906. doi:10.1128/AAC.00906-10.
23. Lam E, Parkin NT (2003) Amprenavir resistance imparted by the I50V mutation in HIV-1 protease can be suppressed by the N88S mutation. Clin Infect Dis 37: 1273–1274. doi:10.1086/378894.
24. Diallo K, Brenner B, Oliveira M, Moisi D, Detorio M, et al. (2003) The M184V substitution in human immunodeficiency virus type 1 reverse transcriptase delays the development of resistance to amprenavir and efavirenz in subtype B and C clinical isolates. Antimicrob Agents Chemother 47: 2376–2379.
25. Marcelin A-G, Affolabi D, Lamotte C, Mohand HA, Delaugerre C, et al. (2004) Resistance profiles observed in virological failures after 24 weeks of amprenavir/ritonavir containing regimen in protease inhibitor experienced patients. J Med Virol 74: 16–20. doi:10.1002/jmv.20140.
26. De Mendoza C, Valer L, Bacheler L, Pattery T, Corral A, et al. (2006) Prevalence of the HIV-1 protease mutation I47A in clinical practice and association with lopinavir resistance. AIDS 20: 1071–1074. doi:10.1097/01.aids.0000222084.44411.cc.
27. Parker MT, Jevons MP (1964) A Survey of Methicillin Resistance in Staphylococcus Aureus. Postgrad Med J 40: SUPPL:170–178.
28. Balla S, Thapar V, Verma S, Luong T, Faghri T, et al. (2006) Minimotif Miner: a tool for investigating protein function. Nature Methods 3: 175–177.
29. Sargeant D, Deverasatty S, Luo Y, Baleta AV, Zobrist S, et al. (2011) HIVToolbox, an integrated web application for investigating HIV. PloS One 6: e20122.
30. Vyas J, Gryk MR, Schiller MR (2009) VENN, a tool for titrating sequence conservation onto protein structures. Nucleic Acids Res 37: e124. doi:10.1093/nar/gkp616.
31. Schmidt B, Korn K, Moschik B, Paatz C, Uberla K, et al. (2000) Low level of cross-resistance to amprenavir (141W94) in samples from patients pretreated with other protease inhibitors. Antimicrob Agents Chemother 44: 3213–3216.
32. Masquelier B, Costagliola D, Schmuck A, Cottalorda J, Schneider V, et al. (2005) Prevalence of complete resistance to at least two classes of antiretroviral drugs in treated HIV-1-infected patients: a French nationwide study. J Med Virol 76: 441–446. doi:10.1002/jmv.20380.
33. Arvieux C, Tribut O (2005) Amprenavir or fosamprenavir plus ritonavir in HIV infection: pharmacology, efficacy and tolerability profile. Drugs 65: 633–659.
34. Pillay D (2004) Current patterns in the epidemiology of primary HIV drug resistance in North America and Europe. Antivir Ther (Lond) 9: 695–702.
35. Bandaranayake RM, Kolli M, King NM, Nalivaika EA, Heroux A, et al. (2010) The effect of clade-specific sequence polymorphisms on HIV-1 protease activity and inhibitor resistance pathways. J Virol 84: 9995–10003. doi:10.1128/JVI.00505-10.
36. Chan PA, Kantor R (2009) Transmitted drug resistance in nonsubtype B HIV-1 infection. HIV Ther 3: 447–465. doi:10.2217/hiv.09.30.
37. Menzo S, Castagna A, Monachetti A, Hasson H, Danise A, et al. (2004) Genotype and Phenotype Patterns of Human Immunodeficiency Virus Type 1 Resistance to Enfuvirtide during Long-Term Treatment. Antimicrobial Agents and Chemotherapy 48: 3253–3259. doi:10.1128/AAC.48.9.3253-3259.2004.

38. Rahim S, Fredrick LM, da Silva BA, Bernstein B, King MS (2009) Geographic and temporal trends of transmitted HIV-1 drug resistance among antiretroviral-naïve subjects screening for two clinical trials in North America and Western Europe. HIV Clin Trials 10: 94–103. doi:10.1310/hct1002-94.

39. Brownstein JS, Freifeld CC, Chan EH, Keller M, Sonricker AL, et al. (2010) Information technology and global surveillance of cases of 2009 H1N1 influenza. N Engl J Med 362: 1731–1735. doi:10.1056/NEJMsr1002707.

40. Janies DA, Treseder T, Alexandrov B, Habib F, Chen JJ, et al. (2011) The Supramap project: linking pathogen genomes with geography to fight emergent infectious diseases. Cladistics 27: 61–66. doi:10.1111/j.1096-0031.2010.00314.x.

8