

Software

Isolation by distance, web service

Jeffrey L Jensen¹, Andrew J Bohonak² and Scott T Kelley^{*2}Address: ¹High Tech High, 2861 Womble Road, San Diego, CA 92106-6025, USA and ²Department of Biology, 5500 Campanile Drive, San Diego State University, San Diego, CA 92182-4164, USAEmail: Jeffrey L Jensen - jljensen@stanford.edu; Andrew J Bohonak - bohonak@sciences.sdsu.edu; Scott T Kelley* - skelley@sciences.sdsu.edu

* Corresponding author

Published: 11 March 2005

Received: 31 October 2004

BMC Genetics 2005, 6:13 doi:10.1186/1471-2156-6-13

Accepted: 11 March 2005

This article is available from: <http://www.biomedcentral.com/1471-2156/6/13>

© 2005 Jensen et al; licensee BioMed Central Ltd.

This is an Open Access article distributed under the terms of the Creative Commons Attribution License (<http://creativecommons.org/licenses/by/2.0>), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

Abstract

Background: The population genetic pattern known as "isolation by distance" results from spatially limited gene flow and is a commonly observed phenomenon in natural populations. However, few software programs exist for estimating the degree of isolation by distance among populations, and they tend not to be user-friendly.

Results: We have created Isolation by Distance Web Service (IBDWS) a user-friendly web interface for determining patterns of isolation by distance. Using this site, population geneticists can perform a variety of powerful statistical tests including Mantel tests, Reduced Major Axis (RMA) regression analysis, as well as calculate F_{ST} between all pairs of populations and perform basic summary statistics (e.g., heterozygosity). All statistical results, including publication-quality scatter plots in Postscript format, are returned rapidly to the user and can be easily downloaded.

Conclusion: IBDWS population genetics analysis software is hosted at <http://phage.sdsu.edu/~jensen/> and documentation is available at <http://www.bio.sdsu.edu/pub/andy/IBD.html>. The source code has been made available on Source Forge at <http://sourceforge.net/projects/ibdws/>.

Background

The term "isolation by distance" (IBD) was first used by Sewall Wright [1,2] to describe patterns of population genetic variation that derive from spatially limited gene flow. IBD is defined as a decrease in the genetic similarity among populations as the geographic distance between them increases. Statistical tests for IBD can be conducted using populations or individuals as the units of replication, although analyses at the individual level typically utilize spatial autocorrelation statistics [3]. When individuals can be grouped into populations, it is possible to calculate 1) a matrix of genetic similarity or distance between all population pairs for comparison with 2) a matrix of pairwise geographic distances. A nonparametric Mantel test is typically used to test for non-random associations

between these two matrices. Bohonak [4] has also suggested performing statistical tests of the IBD slope and intercept using bootstrapped pseudoreplicates, and recently, Yang [5] has developed a likelihood-based approach for IBD analysis. The ever-growing volume of population genetic data and a continued desire to interpret spatial processes make computational tools for rapidly identifying patterns of isolation by distance highly desirable.

In this paper we describe a sophisticated, yet user-friendly, Common Gateway Interface (CGI) implementation of statistical software for detecting IBD patterns. Isolation by Distance Web Service (IBDWS) is a major expansion and upgrade of Isolation By Distance (IBD), a standalone

computer program written in C and originally compiled for Windows and Macintosh [4]. IBDWS includes the following highly useful features for analysis of genetic structure in natural populations:

- (1) Analysis of raw data: For a raw data set with codominant markers (e.g., microsatellites), IBDWS calculates basic summary statistics including heterozygosity, F_{ST} between all pairs of populations, and standard transformations of F_{ST} .
- (2) Statistical tests: IBDWS performs Mantel tests, partial Mantel tests, reduced major axis (RMA) regression of IBD slope and intercept, and error estimation for slope and intercept using multiple methods.
- (3) Simple form interface: Various parameters (e.g., number of randomizations) are easily configured on the IBDWS front page and are checked for validity.
- (4) Analysis selection: Multiple analyses may be performed via checkbox selections.
- (5) Publication quality graphics: IBD scatter plots are generated in Postscript format for download and further modification.
- (6) Rapid analysis: IBDWS is hosted on a fast server with minimal wait times (see Results and discussion).

The majority of these features were not included in IBD. Additionally, IBDWS includes thirty-one updates and bug fixes to the original IBD statistical package. Altogether, these changes make IBDWS a major advancement on the original release in terms of both performance and utility.

Implementation

IBDWS was written in C++, compiled with the Unix g++ compiler, and runs on a server using Apache 2.0.50 and Fedora Linux. Postscript file plotting capability (see below) was implemented using CGraph, a C++ plotting library. The UNIX command "convert" is called by IBDWS to create a JPEG for web-based viewing from the postscript file. Both files are held on the server for 24 hours until routine deletion.

Depending on the specific configuration options, IBDWS performs four primary functions. The first is generating standard summary statistics from raw data (if provided). These include a matrix of the genetic distance F_{ST} [1] between all pairs of populations, and several standard transformations of this statistic [see [6,7]]. Second, the scatter plot of genetic vs. geographic distance is generated in JPEG format (for viewing on the web page) and in vector-based Postscript format (for downloading and manip-

ulating in a graphics design program). The Reduced major axis regression line overlays the plot (see Figure 1). An algorithm that evaluates the range of values on each axis chooses an appropriate set of evenly spaced units for labelling.

Third, IBDWS assesses whether there is a statistically significant relationship between the genetic distance (or similarity) matrix and the comparable matrix of geographic distances. As in other software applications, a Mantel test [8] is used to test this relationship, as it properly accounts for the population as the unit of replication, rather than population pairs (i.e., the matrix entries). In addition, IBDWS calculates the slope and intercept of the IBD relationship using reduced major axis (RMA) regression. RMA regression is more appropriate than standard linear regression when both the dependent and the independent variables are measured with error. Error for the slope and intercept are estimated using standard linear approximations, using a one-delete jackknife across population pairs, and by bootstrapping pseudoreplicates over *independent* population pairs (see [4] and the online IBD manual). The final two methods are preferred, as they correctly treat the population as the unit of replication. Figure 2 provides examples of IBDWS statistical output.

Different researchers have recommended logarithmic (base 10) transformation of genetic distance, geographic distance or both, prior to IBD analysis [6,9]. IBDWS automatically performs each of these transformations and provides an interface that switches among the four possible scatter plots.

Results and discussion

Input

IBDWS analyzes genetic data with or without corresponding geographical information. Without geographic distances, summary statistics can be calculated from raw data sets that include diploid genotypes at one or more loci. These statistics include heterozygosity, and estimates of genetic similarity between all population pairs. The additional input of geographic distance between all population pairs facilitates the variety of IBD calculations described above. The geographic distance matrix can be calculated using simple Euclidean distances, or with more complex algorithms that follow biologically relevant landscape features. Input data can be uploaded as an ASCII text file or the data can be inserted into an HTML form on the IBDWS web page (Figure 3). White space delimiters are used between fields. If the population genetic distance matrix is entered directly (rather than individual-level data), the input file may be in matrix or list format (examples of each are provided on the IBDWS home page). Indicator labels (e.g., GENETIC_DISTANCE, GEOGRAPHIC_DISTANCE) are required to separate

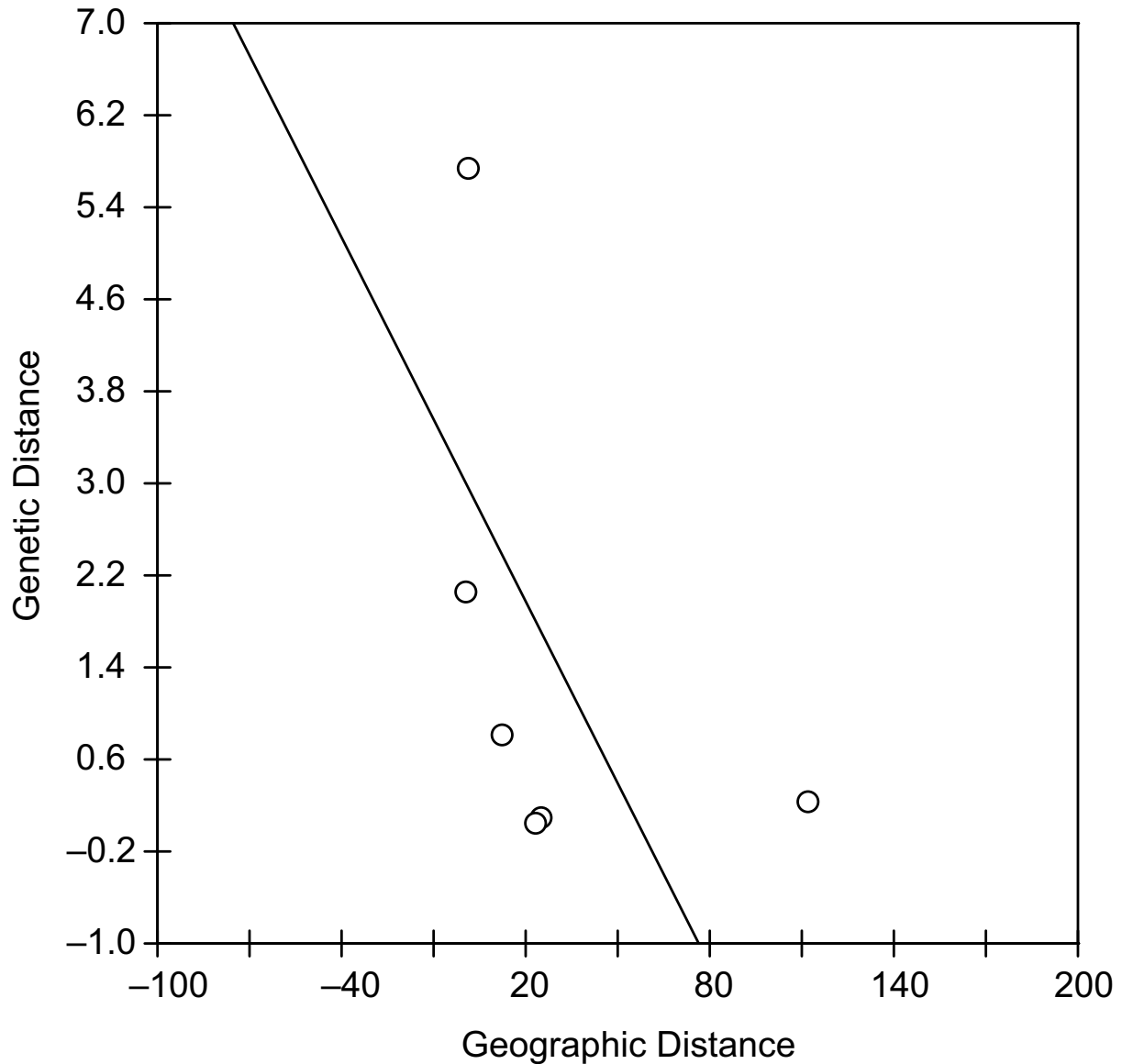


Figure 1
Example scatter plot. The plot was generated using the example raw allelic data and geographic information from the IBD.2 example data set available on the IBDWS website. The RMA regression line overlays the scatter plot. The plot was downloaded in postscript format from the website and slightly modified using Adobe Illustrator.

blocks in the input file or test string, assuring that it is parsed correctly. Additional documentation regarding input for partial Mantel tests is provided in the IBD manual (available at the web site).

Checkboxes and radio buttons allow the user to configure which analyses will be performed (Figure 3). These include the calculation of Rousset's genetic distance [7] from raw data, options to log-transform genetic and/or

IBDWS Results

Input Data Information:

- The raw data header was detected in the input data.
- There are 3 loci, 4 alleles, and 30 individuals.
- Data successfully read in.

Allele Statistics: Allele counts for each population and locus

Population #	Locus #	Allele 1	Allele 2	Allele 3	Allele 4	H[o]
Population 12	Locus 1	1	2	29	0	0.1875
	Locus 2	0	32	0	0	0.0000
	Locus 3	2	3	26	1	0.1875
						Average: 0.1250
Population 2	Locus 1	0	0	32	0	0.0000
	Locus 2	2	30	0	0	0.1250
	Locus 3	0	1	33	0	0.0588
						Average: 0.0613
Population 3	Locus 1	0	20	38	0	0.1379
	Locus 2	0	56	0	0	0.0000
	Locus 3	20	0	38	0	0.0000
						Average: 0.0460
Population 4	Locus 1	16	3	0	1	0.4000
	Locus 2	0	19	1	0	0.1000
	Locus 3	19	1	0	0	0.1000
						Average: 0.2000

F-Statistics

Population	Population	Locus	Fst	M
12	2	1	0.04444	
12	2	2	0.03333	
12	2	3	0.04307	
12	2	Total:	0.04174	5.73984
12	3	1	0.13669	
12	3	2	Undefined	
12	3	3	0.08245	
12	3	Total:	0.10837	2.05685
12	4	1	0.74218	
12	4	2	0.02509	
12	4	3	0.73956	
12	4	Total:	0.72655	0.09409
2	3	1	0.26013	
2	3	2	0.06213	
2	3	3	0.22971	
2	3	Total:	0.23494	0.81412
2	4	1	0.86158	
2	4	2	-0.01079	
2	4	3	0.92573	
2	4	Total:	0.84474	0.04595
3	4	1	0.54805	
3	4	2	0.06056	
3	4	3	0.49686	
3	4	Total:	0.51832	0.23233

Figure 2
IBDWS example statistical results output. The figure shows a partial list of statistical output calculated from the IBD.2 sample data set. Results of Mantel tests and other calculations are also provided. All results can be downloaded as text files.

geographic distances, replacement values for undefined transformations, and the number of bootstrap randomizations per analysis (up to 30,000).

Output

A full analysis of raw data for six loci, two to three alleles per locus, and 326 individuals required an average wait time of 10 seconds. When an input without any geographic distances is used, wait times are negligible. If raw genotypic data are supplied, IBDWS returns an HTML table with basic summary statistics, the genetic distance F_{ST} between all population pairs, the genetic similarity statistic M [6] between all population pairs, and Rousset's [7]

distance $F_{ST} / (1 - F_{ST})$ if desired. If geographic distances are included, IBDWS will display a scatter plot of genetic distance vs. geographic distance for all pairs of populations, and overlay the RMA regression line (Figure 1). If log-transformation of the data is selected, a pull-down menu below the plot allows the user to choose data sets with one or both axes log-transformed. Clicking on any plot downloads a Postscript version to the user's computer. Statistical analyses of the IBD relationship, including Mantel tests, partial Mantel tests (if appropriate) and the RMA regression analyses described above, are summarized in HTML tables as well. A summary of all analyses can be downloaded as a static HTML file, or a tab-delimited

1. INPUT DATA

Input Data Filename:

OR:

File Data Here

2. BASIC SETTINGS

(Raw data format only): If $F \leq 0$, what value should be assigned to Slatkin's M for RMA regression (Should be a large number)?

For IBD plots, would you like a background grid?

Number of randomizations (1-30,000):

Also perform analysis on Logarithmic (Genetic Distance)?

Also perform analysis on Logarithmic (Geographic Distance)?

(Raw data format only): Calculate Rousset's distance?

Calculate RMA regression residuals?

3. DETAILED SETTINGS

(Raw data format only): Calculate summary stats for each locus?

For first set of analyses, use negative genetic distances? (Otherwise, they will be set to zero.)

When log-transforming (genetic distance), set genetic distances less than zero to:

When log-transforming (genetic distance), negative genetic distances will be: Set to value above Thrown out

If calculating Rousset's distance, set genetic distances of 1.00 to:

4. SUBMIT

Figure 3
IBDWS forms interface. The figure shows a portion of the CGI web interface. Data can be uploaded from a text file using the Browse button or pasted directly into the form. Below the data input portion is a partial list of checkboxes for tailoring analyses. A total of twelve parameters are modifiable by the user.

text file suitable for importation into a spreadsheet application.

Future developments

At present, IBDWS requires the user to calculate all pairwise genetic distances between populations for DNA data sets. In future versions, we plan to include algorithms for calculating genetic distances using both sequence and allelic data, including genetic distances based on phylogenetic trees under different models of evolution. When implemented, the user will simply have to supply the raw data and geographic distances before calculations are performed. IBDWS will also be updated with new population genetics analyses as they become available.

Conclusion

In summary, IBDWS is a substantial expansion of the previous PC-based versions of IBD, providing geneticists with a simple interface, multiple statistical calculations, rapid analysis, and high-quality graphics. It will be useful to a significant number of molecular ecology, conservation genetics, population genetics and genomics researchers, and will be updated with newer analyses in the future.

Availability and requirements

IBDWS is currently hosted at <http://phage.sdsu.edu/~jensen/>. Documentation of the statistical calculations and biological context are available on IBD's original website at <http://www.bio.sdsu.edu/pub/andy/IBD.html>. The programming source files have been made available at <http://sourceforge.net/projects/ibdws/> under the GPL (GNU Public License).

List of abbreviations used

IBD: Isolation by distance.

IBDWS: Isolation by distance web service.

RMA: Reduce major axis (regression).

Authors' contributions

JLJ designed and implemented IBDWS. This effort included creating the CGI, debugging and enhancing IBD, implementing postscript graphing libraries and extensive testing of the software. JLJ also wrote the initial draft of the manuscript. AJB wrote the original IBD software, the precursor to IBDWS. AJB also upgraded the code, added graphing algorithms, provided technical advice and assistance to JLJ and contributed to the writing of the manuscript. STK was involved in the early training of JLJ for the project as a mentor in the High Tech High Internship program. STK also guided the development of the project, provided technical advice, and contributed substantially to the writing of the manuscript. All authors read and approved the final manuscript.

Acknowledgements

The authors are grateful to Forest Rohwer for hosting IBDWS on his web server. We also thank Pat McNairnie for helping us with web configuration, account management and troubleshooting tips. We thank Varykina Thackray for editorial suggestions on the manuscript. Finally, we thank Jay Vavra, Stacey Praizner, and members of the High Tech High School Biotechnology program for arranging the internship that lead to this research project.

References

1. Wright S: **Size of population and breeding structure in relation to evolution.** *Science* 1938, **87**:430-431.
2. Wright S: **Breeding structure of populations in relation to speciation.** *American Naturalist* 1940, **74**:232-248.
3. Epperson BK, Li T: **Measurement of genetic structure within populations using Moran's spatial autocorrelation statistics.** *Proc Natl Acad Sci U S A* 1996, **93**:10528-10532.
4. Bohonak AJ: **IBD (Isolation By Distance): a program for analyses of isolation by distance.** *J Hered* 2002, **93**:153-154.
5. Yang RC: **A likelihood-based approach to estimating and testing for isolation by distance.** *Evolution* 2004, **58**:1839-1845.
6. Slatkin M: **Isolation by distance in equilibrium and non-equilibrium populations.** *Evolution* 1993, **47**:264-279.
7. Rousset F: **Genetic differentiation and estimation of gene flow from F-statistics under isolation by distance.** *Genetics* 1997, **145**:1219-1228.
8. Manly BFJ: *Multivariate statistical methods: a primer* 2nd edition. New York: Chapman and Hall; 1994.
9. Hutchinson DW, Templeton AR: **Correlation of pairwise genetic and geographic distance measures: inferring the relative influences of gene flow and drift on the distribution of genetic variability.** *Evolution* 1999, **53**:1898-1914.

Publish with **BioMed Central** and every scientist can read your work free of charge

"BioMed Central will be the most significant development for disseminating the results of biomedical research in our lifetime."

Sir Paul Nurse, Cancer Research UK

Your research papers will be:

- available free of charge to the entire biomedical community
- peer reviewed and published immediately upon acceptance
- cited in PubMed and archived on PubMed Central
- yours — you keep the copyright

Submit your manuscript here:
http://www.biomedcentral.com/info/publishing_adv.asp

