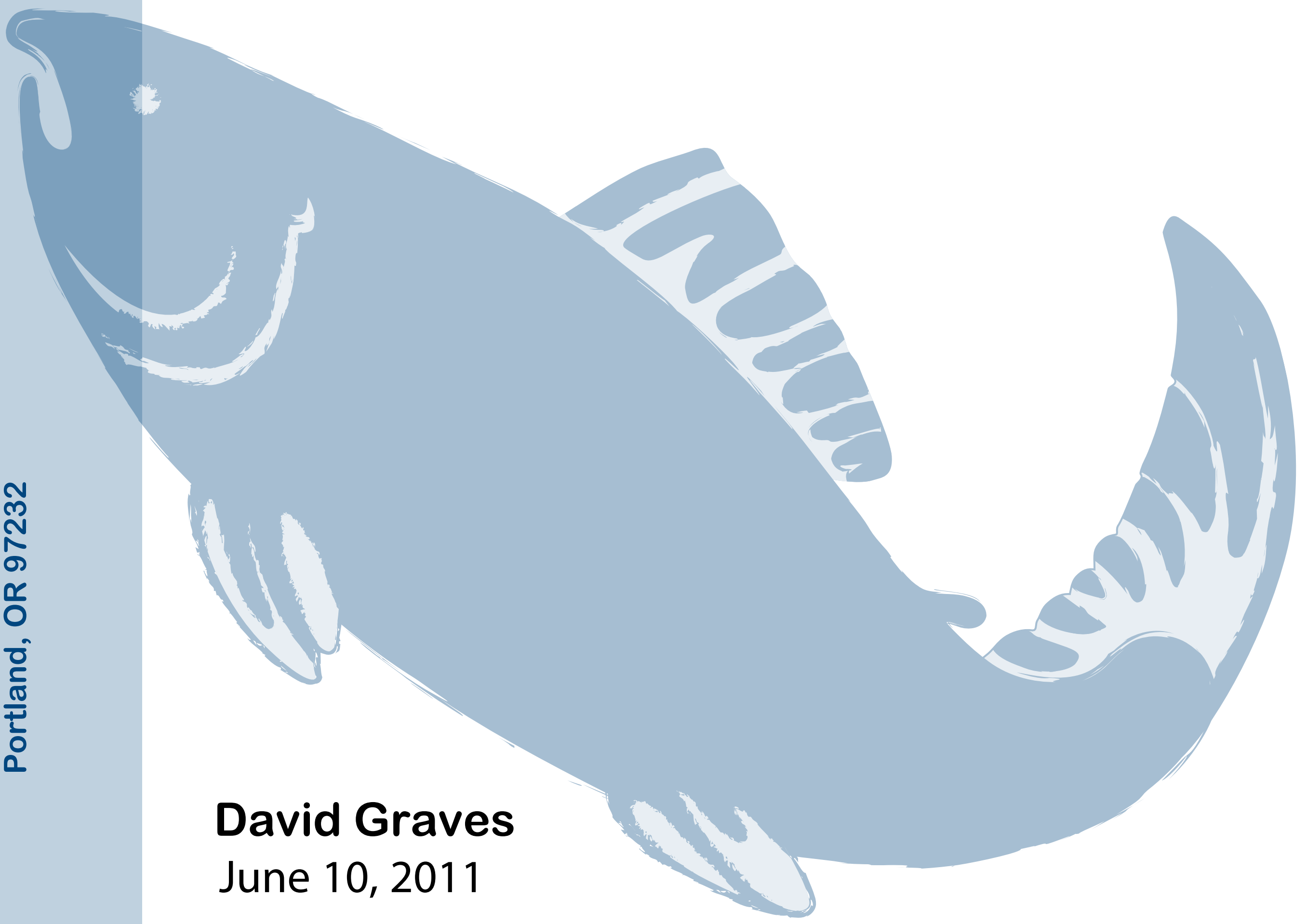
CRITFC

TECHNICAL REPORT 11-03

A Computer Program for Calculating the Distance Between Sites in the Columbia River Basin



David Graves

June 10, 2011

Columbia River Inter-Tribal Fish Commission
729 NE Oregon, Suite 200
Portland, OR 97232
503.238.0667
www.critfc.org

A computer program for calculating the distance between sites in the Columbia River basin

David Graves, The Columbia River InterTribal Fish Commission

June 10, 2011

Abstract

In landscape genetics analyses, it is useful to measure the physical distance between collections or populations of interest in order to study the effect of distance on genetic diversity and population structure. Simple measurements such as Euclidean distance can be misleading, however, because fish and other species may be restricted to travel through relatively circuitous migration pathways such as a stream network rather than directly overland across the landscape. The Pairwise Stream Distance Application is a tool developed in Microsoft Access software to resolve the fluvial distance of the shortest path between pairs of sites in a stream network. It may be operated on a single pair of sites or simultaneously on a batch of multiple sites to generate a pairwise distance matrix. The tool is built with comprehensive stream information for the Columbia River Basin of the U.S., but may also be modified for use in other river basins. It is available for free download and use from a web address at the Columbia River Inter-Tribal Fish Commission.

Body

Landscape genetics integrates the disciplines of population genetics, landscape ecology, and spatial analysis to examine the roles that landscape variables have on genetic diversity and population structure of organisms (Storfer *et al*. 2007). For studies of landscape and its effect on the genetic and life history attributes of terrestrial and aquatic organisms, physical distance is often an important measure. Isolation-by-distance gene-flow is a process that can be inferred when genetic differentiation between individuals or populations is found to correlate with physical distances that separate them (Manel *et al*. 2003). In the example of anadromous fish species such as salmon and steelhead, fluvial distance between the spawning locations of populations is a significant factor in their genetic structure and life history diversity, although other landscape factors also can be important (Hendry *et al*. 2000; Primmer *et al*. 2006; Narum *et al*. 2008). In the Pacific Northwest, groups of salmon that reside in proximate locations throughout their fresh water life history tend to be more genetically similar, while those that are more isolated exhibit greater genetic differentiation (Waples 1991)

Spatial autocorrelation occurs where samples that are physically proximate to one another share similar traits (Rogerson 2001; Fortin and Dale 2005). There are different spatial methods to measure physical distance between individuals and populations. The simplest approach is to measure the Euclidean ("as the crow flies") distance overland between sites where populations have been sampled. This method

may be suitable in some landscape genetic applications depending on the behavior of the species. For example, in a hypothetical case where terrestrial organisms are able to migrate freely across a landscape, overland distance may represent an accurate measure of physical distance between individuals. But a Euclidean distance measurement may produce poor results where a species is restricted to certain migration pathways. For example, two populations of fish may occupy adjacent watersheds separated by a ridgeline. An overland distance measurement may show them to be proximate to each other, but because fish cannot migrate where a navigable stream or water body is absent, this measurement is misleading. A more accurate measurement of intermediate or pairwise distance must reflect migration through available pathways, which in this case is the fluvial distance through the stream network between two sites.

In order to assist with landscape genetics studies, we have developed a tool (The Pairwise Stream Distance Application) to measure the fluvial distance via a stream network between disparate sites where fish have been sampled. This tool was built in the Microsoft Access database software using the Visual Basic for Applications programming language, and operates solely in this environment as a standalone database. Geographic Information Systems (GIS) measurements of the stream network were used to inform the tool development, but GIS software is not needed to run the tool. A user may operate the tool for two sites at a time to find their closest path and its distance via the stream network, or may run the tool simultaneously on a batch of several coordinates in order to determine the distance and path between all possible combinations in a batch of sites (e.g. for construction of a pairwise matrix of geographic distances). The tool incorporates the stream network for the U.S. portion of the Columbia River Basin, which encompasses 21,070 named streams and rivers, and 32,978 unnamed streams. The tool may also be customized to work with other stream networks.

The foundation of the Pairwise Stream Distance Application is a stream network table (LLIDStream), which lists length and tributary information for all of the streams and rivers in the U.S. portion of the Columbia River Basin. This table was developed by the StreamNet project of the Pacific States Marine Fisheries Commission (http://www.streamnet.org) through GIS analysis of a 1:100,000-scale hydrography layer of the Pacific Northwest. Each stream in the network is described by a 13-digit identifier (referred to as its StreamID or LLID), a feature name, total stream length in kilometers, the identifier of the stream it flows into (referred to as its TribID), and the upstream distance of the StreamID from the mouth of the TribID (referred to as its TribKM). A subbasin attribute was added to this table through a GIS overlay to more easily sort streams into their respective watersheds. This table represents an entire stream network with integrity because all connected streams flow to a single feature (the Pacific Ocean), and all distances and tributary intersections are measured and included in the table.

The tool is programmed to assess the path along the stream network between pairs of stream-based sites. It first tests whether the coordinates for these sites are located upstream/downstream of each other or are in adjacent watersheds of a common downstream river (Figure 1). All possible pairs of coordinates will conform to one of these situations. After this determination, the program evaluates connections through tributaries and streams through the stream network table to find the path between the two sites. As it does this, it accumulates all connecting distances between the two sites

while recording the path in a text string. When the path between the two sites has been resolved, the user is provided with a report of the distance and text string of the path. In the case of a batch operation, it reads multiple site-coordinates from a table in the database (Input_PairwiseBatch), resolves the path and distance between all possible pairs of these coordinates, and writes the results to a different table (Output_PairwiseBatch).

Figure 1. Example of two different spatial relationships between sites

All tool operations are accessed through a simple entry form (Figure 2). To operate the tool, a user must first identify the coordinates of their sites along the stream network. A stream coordinate is composed of the identifier of the stream, and the distance upstream from the mouth of the stream where the site is located. For example, if a fish was sampled at river kilometer 4.20 on the Wenatchee River in

Washington, its stream coordinate is the unique identifier for the Wenatchee River (1203156474560) and the upstream distance (4.20 km). The distance upstream is as measured on a 1:100,000-scale quadrangle map. After a user identifies a second stream coordinate (for example, river kilometer 2.50 on Brock Creek (1222187438746) in Oregon), they may then run the application to determine the distance and path between these two sites via the stream network. The tool prompts the user to enter the two coordinates, and then returns the distance between the two sites (e.g. 989.28 kilometers), and the path between the two sites (e. g. "Wenatchee River -> Columbia River and Brock Creek -> N Fk Middle Fk Willamette R -> Middle Fork Willamette River -> Willamette River -> Columbia River"). If a second stream coordinate is not specified, the application will by default find the distance and path from the input site to the Pacific Ocean at the mouth of the Columbia River (the most downstream point in the river network). When using the batch operation, the user enters a set of stream coordinates into a table before running the program.

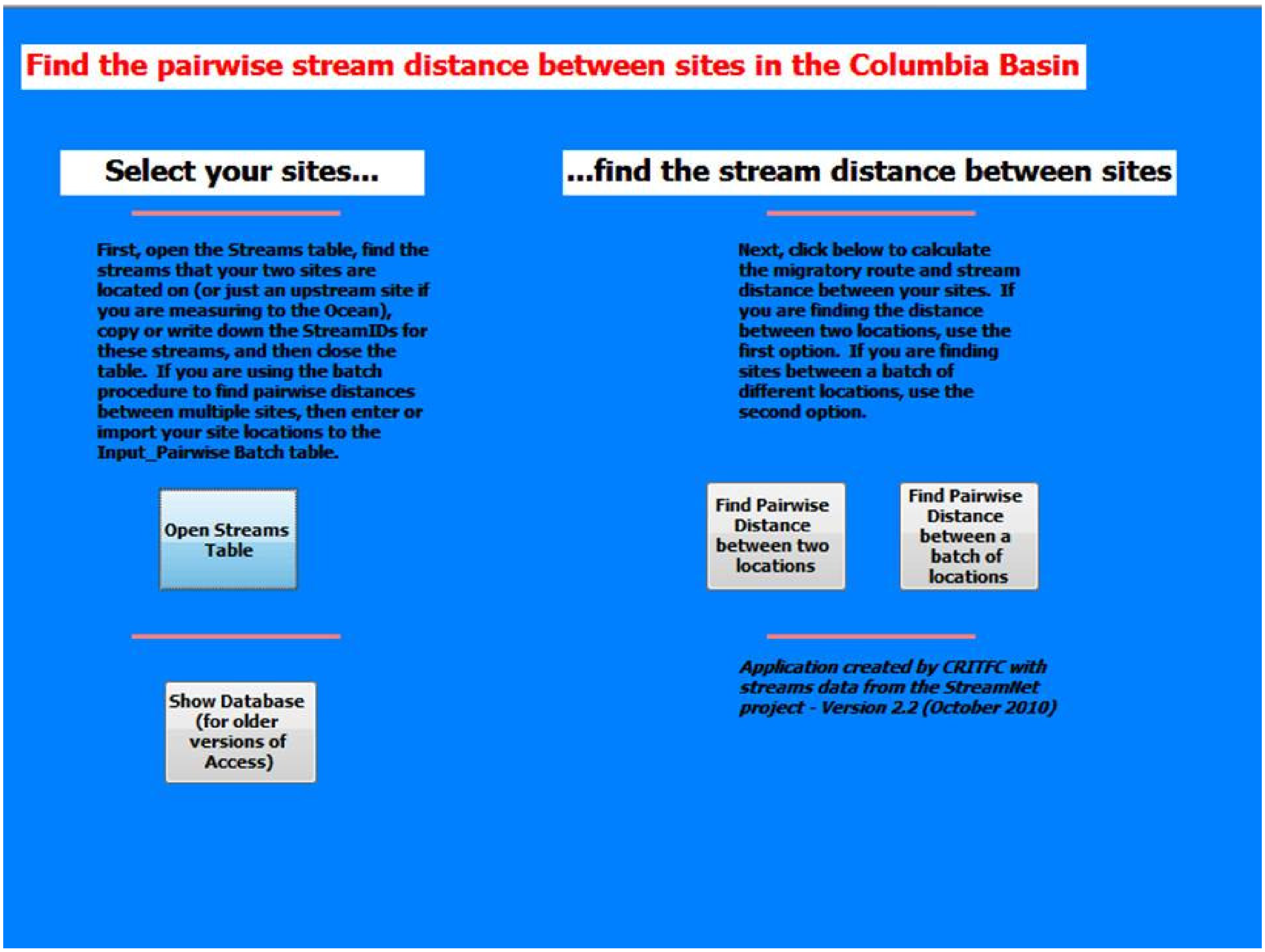


Figure 2. Pairwise Stream Distance Application Entry Form

Instructions are provided with the application for determining a stream coordinate when a geographic XY coordinate (such as longitude and latitude) is known but the stream coordinate is unknown. In this case, a GIS and the stream layer for the Columbia River Basin are needed to determine the correct stream coordinate for each XY coordinate. The tool may also be customized to work with stream networks outside of the Columbia River Basin by entering the information about all streams and their network into the LLIDStream table. Typically, one would use a GIS with a routed streams layer to derive this stream network information in a table format. All fields of the LLIDStream table need to be filled, so it is necessary to script a GIS process that calculates the length of each stream in the network as well as the intersection of streams and the distance upstream where these intersections occur along the network.

The Pairwise Stream Distance Application allows one to quickly compute the distance along a stream network between many different sites. This may be helpful to geneticists and other researchers who wish to assess the biological relevance of physical distance measurements between sample sites, and may further benefit those who already perform this process regularly through slower methods such as manual measurements of distances in a GIS or on paper maps. The tool may also be helpful in other types of analyses such as in the measurement of water pollution pathways. The interface requires Microsoft Access (version 97 or later) to run, but is straightforward to use for anyone with a basic familiarity with this software. The tool and its associated documentation are available as a free download from the Columbia River InterTribal Fish Commission website from the following URL: http://maps.critfc.org/file_download/StreamDistanceApplication.zip.

Acknowledgement

This work benefitted from discussions with A. Matala, J. Hess, and S. Narum at the Columbia River Inter-Tribal Fish Commission, and M. Banach at the Pacific States Marine Fisheries Commission. We also thank V. Hare at the Pacific States Marine Fisheries Commission for providing the stream network table for use in this program.

References

Fortin M. J. and M. Dale. 2005. *Spatial Analysis: A Guide for Ecologists*. Cambridge University Press, Cambridge, UK.

Hendry A. P., J. K. Wenburg, P. Bentzen , E. C. Volk, T. P. Quinn. 2000. Rapid Evolution of Reproductive Isolation in the Wild: Evidence from Introduced Salmon. *Science* 290: 516-518.

Manel S., M. K. Schwartz, G. Luikart, P. Taberlet. 2003. Landscape genetics: combining landscape ecology and population genetics. *Trends in Ecology and Evolution* 18: 189-207.

Narum S. R., J. S. Zendt, D. Graves, W. S. Sharp. 2008. Influence of landscape on resident and anadromous life history types of Oncorhynchus mykiss. *Canadian Journal of Fisheries Aquatic Sciences* 65: 1013-1023.

Primmer C. R., A. J. Veselov, A. Zubchenko, A. Poututkin, I. Bakhmet, M. T. Koskinen. 2006. Isolation by distance within a river system: genetic population structuring of Atlantic salmon, *Salmo salar*, in tributaries of the Varzuga River in northwest Russia

Rogerson P. A. 2001. *Spatial Methods for Geography.* Sage Publications, London, UK.

Storfer A., M. A. Murphy, J. S. Evans, C. S. Goldberg, S. Robinson, S. F. Spear, R. Dezzani , E. Delmelle, L. Vierling, L. P. Waits. 2007. Putting the 'landscape' in landscape genetics. *Heredity* 98: 128-142.

Waples R. S. 1991. Definition of "species" under the Endangered Species Act: application to pacific salmon. *National Oceanic and Atmospheric Administration Technical Memorandum NMFS F/NWC-194*.