

Genetics and population analysis

## PedVizApi: a Java API for the interactive, visual analysis of extended pedigrees

Christian Fuchsberger<sup>1,\*</sup>, Mario Falchi<sup>2</sup>, Lukas Forer<sup>1</sup> and Peter P. Pramstaller<sup>1</sup>

<sup>1</sup>Institute of Genetic Medicine, European Academy, Bolzano, Italy and <sup>2</sup>Twin Research & Genetic Epidemiology Unit, Kings College London, London, UK

Received on September 11, 2007; revised on November 1, 2007; accepted on November 17, 2007

Advance Access publication November 22, 2007

Associate Editor: Keith Crandall

### ABSTRACT

**Summary:** PedVizApi is a Java API (application program interface) for the visual analysis of large and complex pedigrees. It provides all the necessary functionality for the interactive exploration of extended genealogies. While available packages are mostly focused on a static representation or cannot be added to an existing application, PedVizApi is a highly flexible open source library for the efficient construction of visual-based applications for the analysis of family data. An extensive demo application and a R interface is provided.

**Availability:** <http://www.pedvizapi.org>

**Contact:** christian.fuchsberger@eurac.edu

Pedigree visualization is a fundamental task in family based studies. With the increased complexity of genealogies, especially those found in large inbred populations, this task becomes quite complex. Researchers can draw large genealogies by using programs, such as Pedigraph, Pedfiddler, CraneFoot (Mäkinen *et al.*, 2005) and Madeline 2.0 PDE (Trager *et al.*, 2007), whereby the analyses become disconnected from the visualization process. Using packages like PedNavigator (Mancosu *et al.*, 2005), PVin and Pedigree Viewer pedigrees can be manipulated in an interactive way, but they focus only on sub-pedigrees, require special hardware or provide only limited interactivity. Furthermore, the integration in popular tools is of high interest (Zhao, 2006). More information is available on the PedVizApi website.

Our contribution is to provide a highly flexible library for the visualization of whole extended pedigrees (>5000 individuals) for building more sophisticated interactive applications by following the principle of 'Visual Analytics' (Thomas *et al.*, 2006).

The challenge of Visual Analytics is to combine the outstanding visual capabilities of humans with the power of analytical methods to support the knowledge discovery process. Most importantly, the user is not only an interpreter of visual and analytical output, but takes an active role in driving the whole process.

For the visualization of whole large pedigrees, we provide 2D and 2.5D layouts. While the 2D layout corresponds to the standard representation of pedigrees, we have developed a novel 2.5D visualization to improve the aesthetics of huge genealogies by preserving their comprehensibility. In this drawing, nodes are distributed on two distinct layers in the 3D space (Fig. 1). Furthermore, nodes can be distributed on the two layers according to different criteria, like disease status, to facilitate the identification of patterns (Kaufmann and Wagner, 2001). Both layout algorithms are highly optimized; however, the 2.5D layout requires a proper graphics-card.

PedVizApi is a versatile tool for mapping different types of data on the genealogy. For qualitative traits a set of standard symbols is available. Since there is an increasing interest on quantitative traits, our visualization uses a novel approach to map this type of data on a pedigree structure without discretization. We use the maximum, minimum and mean values to calculate an appropriate colour bar. Then, around the individual's value, we extract a subsection of the bar and map this on the genealogy.

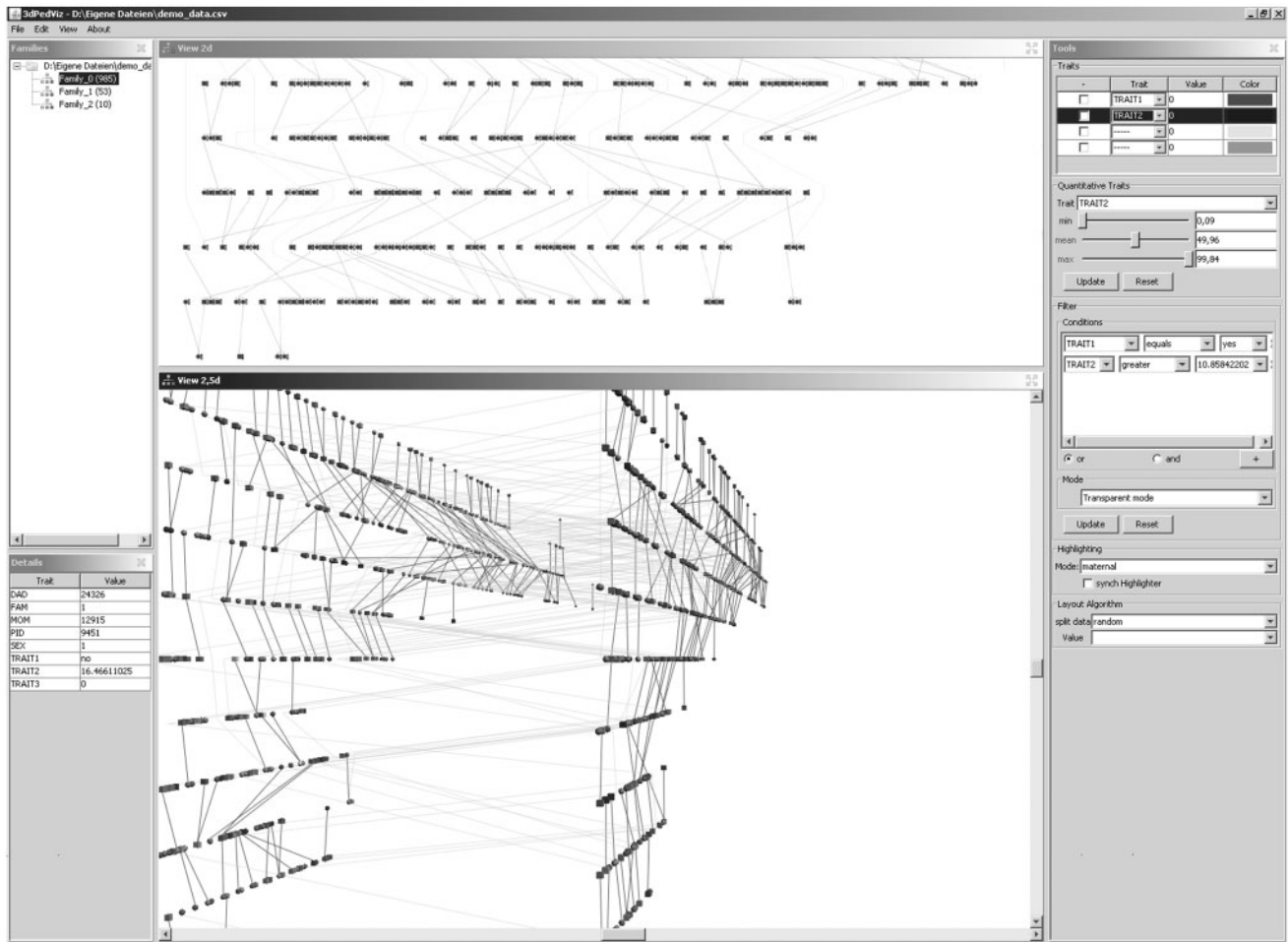
Depending on the research question different additional information, such as detailed phenotypes, is needed and has to be included in the pedigree drawing. PedVizApi supports different data sources, such as relational databases and PED files to provide these details on demand. The extension of PedVizApi to genotype data can be done easily.

To support the user in focusing on the 'important', a set of interactions are integrated. Different paths, like maternal and paternal lineages, all ancestors, all successors, shortest path between two individuals, can be highlighted. Since only limited information is available for older generations, they can be visualized in a compact way. Furthermore, the integrated fisheye-view allows the user to focus on details while conserving context (Preece *et al.*, 2002).

The identification of interesting subsets can be done by dynamic filtering as provided by the demo application (Fig. 1). The user can formulate different complex filters like, e.g., *trait1 = yes AND trait2 > 10.0*, and get continuous feedback about the selected subset by blinking the corresponding individuals or fading the deselected elements (Shneiderman, 1994).

Various types of view transformation support the user during the explorative process. Zoom, move, rotate and incline are

\*To whom correspondence should be addressed.



**Fig. 1.** PedVizApi demo application: the interface consists of three columns. A 1270 individuals pedigree drawn in 2D and in 2.5D is shown in the main window. The left column consists of the family tree (if more families are available in the data source, they will be identified and provided) and the detail-on-demand panel in the lower part. In the right column (from top to bottom): two panels for the mapping of trait/colours and the mapping of quantitative traits, a panel for building dynamic filters, one for the selection of different highlighting methods and a panel for choosing the node distribution mode for the 2.5D layout.

basic technologies. More advanced concepts are the marking of single individuals of interest in the whole genealogy to extract the corresponding sub-pedigree. These sub-pedigrees can be visualized in a number of windows for simultaneous comparison. Connections are conserved and individuals are highlighted with the same colour in the different windows.

Integrating PedVizApi into other applications is quite straightforward and the basic framework is added as follows:

```
final Graph g = new Graph();
DBGraphLoader loader =
    new DBGraphLoader(db, table);
loader.load(g);
Sugiyama sy = new Sugiyama(g); sy.run();
GraphView3D gView =
    new GraphView3D(sy.getLayoutedGraph());
```

This integration is not limited to Java-based programs, as demonstrated by the provided R interface. In addition,

PedVizApi can be extended simply on two different levels: first, new rules for modifying nodes and edges can be added. Second, as the library is programmed in a strictly object-orientated way new objects can be derived easily. Furthermore, PedVizApi has been used successfully for exploration and subsequent epidemiological analysis of Microisolates (Pattaro *et al.*, 2007). Finally, Jenti, a tool for mining complex inbred genealogies (in this issue of the journal), integrates PedVizApi into its analysis process and underlines the value of such a visual approach.

## ACKNOWLEDGEMENTS

This work was supported by the Ministry of Health of the Autonomous Province of Bolzano and the South Tyrolean Sparkasse Foundation.

*Conflict of Interest:* none declared.

---

**REFERENCES**

- Kaufmann,M. and Wagner,D. (2001) Drawing graphs: methods and models. In *Lecture Notes in Computer Science*. Vol. 2025, Springer-Verlag, New York.
- Mäkinen,V.P. *et al.* (2005) High-throughput pedigree drawing. *Eur. J. Hum. Genet.*, **13**, 987–989.
- Mancosu,G. *et al.* (2005) Browsing isolated population data. *BMC Bioinformatics*, **6** (Suppl. 4).
- Pattaro,C. *et al.* (2007) The genetic study of three population microisolates in South Tyrol (MICROS): study design and epidemiological perspectives. *BMC Med. Genet.*, **8**, 29.
- Preece,J. *et al.* (2002) Interaction Design. John Wiley & Sons, New York, USA.
- Shneiderman,B. (1994) Dynamic queries for visual information seeking. *IEEE Softw.*, **11**, 70–77.
- Thomas,J.J. *et al.* (2006) A visual analytics agenda. *IEEE Comput. Graph. Appl.*, **26**, 10–13.
- Trager,E.H. *et al.* (2007) Madeline 2.0 PDE: a new program for local and web-based pedigree drawing. *Bioinformatics*, **23**, 1854–1856.
- Zhao,J.H. (2006) Pedigree-drawing with R and graphviz. *Bioinformatics*, **22**, 1013–1014.