The HuGeMap Database: interconnection and visualization of human genome maps

Emmanuel Barillot[1,2,*], Stuart Pook[1,3], Frédéric Guyon[1], Christophe Cussat-Blanc[1,2], Éric Viara[4] and Guy Vaysseix[1,2]

[1]GIS Infobiogen, 7 rue Guy Môquet BP 8, 94801 Villejuif cedex, France, [2]Généthon, 1 rue de l'Internationale BP 60, 91002 Évry cedex, France, [3]École Nationale Supérieure des Télécommunications, Département Informatique et Réseaux, 46 rue Barrault, 75634 Paris cedex 13, France and [4]Sysra Informatique, 7 rue de Bièvres, 92140 Clamart, France

ABSTRACT

The HuGeMap database stores the major genetic and physical maps of the human genome. HuGeMap is accessible on the Web at http://www.infobiogen.fr/ services/Hugemap and through a CORBA server. A standard genome map data format for the interconnection of genome map databases was defined in collaboration with the EBI. The HuGeMap CORBA server provides this interconnection using the interface definition language IDL. Two graphical user interfaces were developed for the visualization of the HuGeMap data: ZoomMap (http://www.infobiogen.fr/services/ zomit/ZoomMap.html) for navigation by zooming and data transformation via magic lenses, and Mappet-Show (http://www.infobiogen.fr/services/Mappet) for visualizing and comparing maps.

THE HuGeMap DATABASE

Contents

The large scale human genome mapping projects have now reached completion. The HuGeMap database contains the final versions of the maps and some of the mapping data produced. It thus integrates the genetic maps from Généthon (1) and the Cooperative Human Linkage Consortium (2), and the physical maps from CEPH Généthon (3–5) and the Whitehead Institute-MIT (6). It also includes a cytogenetic description of the human genome as described by Harnden and Klinger (7).

Architecture

The HuGeMap database is managed by the Object-Oriented Database Management System EYEDB (http://www.infobiogen.fr/ services/eyedb). The EYEDB system offers: (i) a WWW access to HuGeMap at http://www.infobiogen.fr/services/Hugemap ; and (ii) a CORBA access to HuGeMap (see below).

INTERCONNECTION TO OTHER GENOME MAP DATABASES

The HuGeMap database stores most of the major human genome maps, however, it is not our intention to include all the maps. Nevertheless it would be useful to link the maps in HuGeMap to other maps such as the human genome radiation hybrid maps (8) or the numerous mammalian maps.

In collaboration with the European Bioinformatics Institute, we have adopted a common and standard solution to interconnection problems. Our solution is based on the Common Object Request Broker Architecture (9) which offers a protocol (Internet Inter-ORB Protocol) and a language (Interface Definition Language, IDL) for the exchange of distributed objects. We have defined and implemented a common data format for genome maps, written in IDL, and we propose it to the mapping community as a standard (10).

For the mapping community, adoption of this standard will allow a real interconnection of genome map databases. Any software using this common genome map IDL will be able to query any database implementing the IDL. HuGeMap and RHdb (8) use this IDL, and it has been proposed to the Life Science Research Task Force of the Object Management Group (the goal of this work group is to 'improve communication and interoperability among computational resources in life sciences research'; http://www.omg.org/lsr/).

Access to our HuGeMap CORBA server is possible using the Interoperable Object Reference downloadable from http://www. infobiogen.fr/services/Hugemap/IOR

VISUALIZATION OF GENOME MAPS

We have designed two different interfaces for visualization and navigation in HuGeMap. They are based on new visualization concepts in biology (detailed below) and query HuGeMap through a CORBA server.

*To whom correspondence should be addressed at: GIS Infobiogen, 7 rue Guy Môquet BP 8, 94801 Villejuif cedex, France. Tel: +33 1 49 58 36 82; Fax: +33 1 45 59 52 50; Email: manu@infobiogen.fr
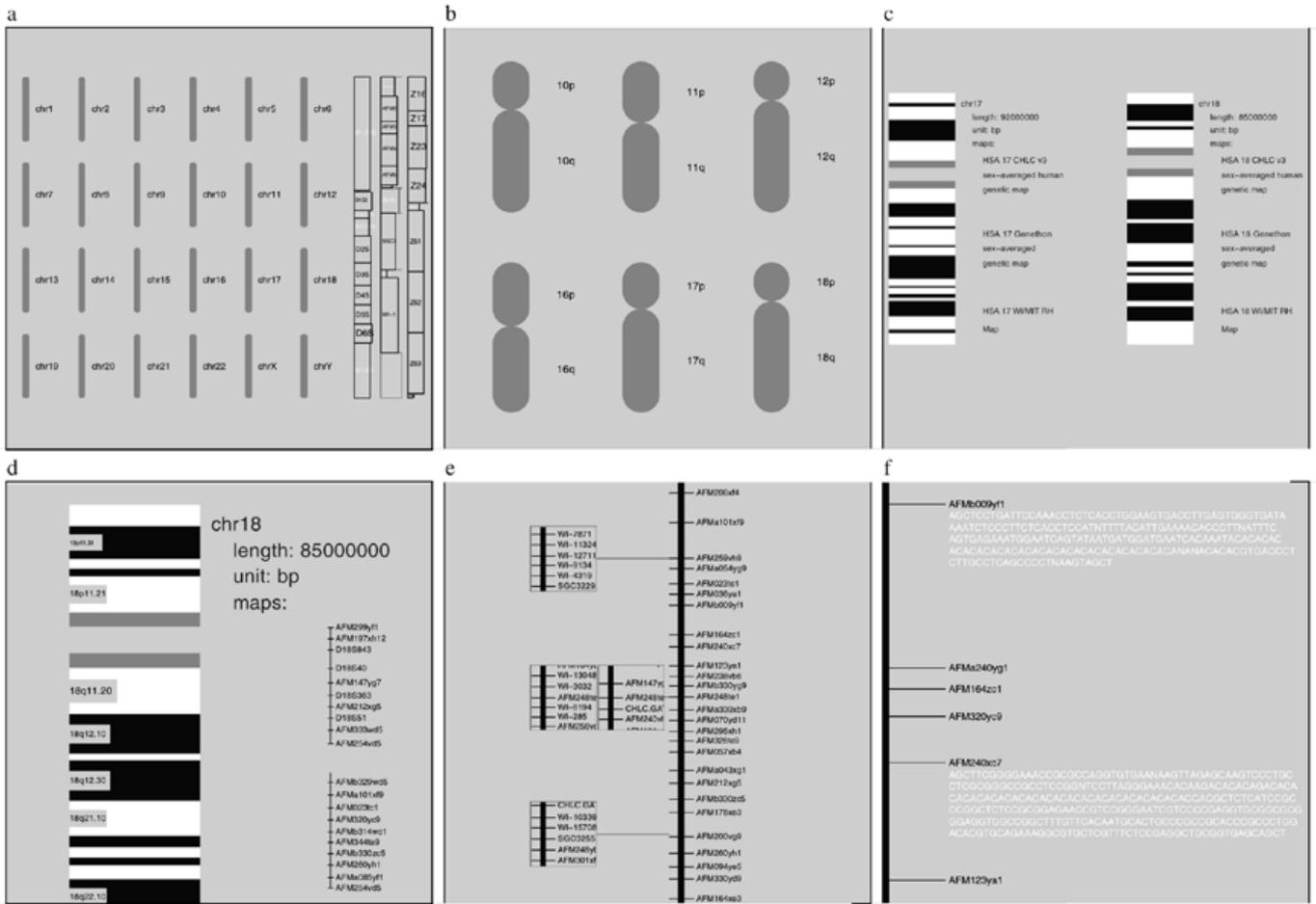
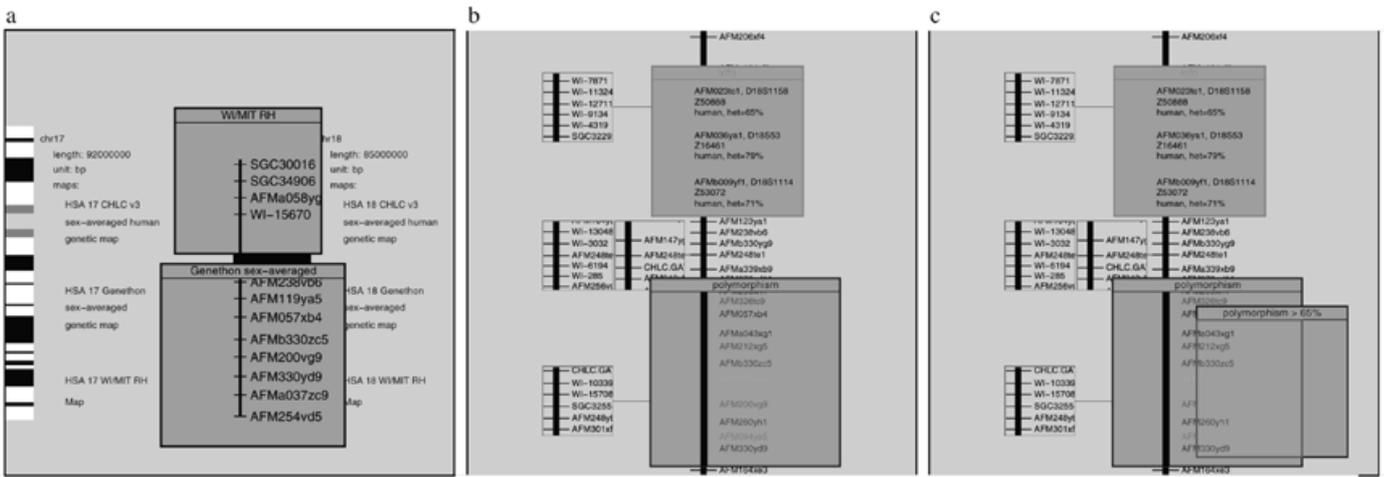**Figure 1.** Zooming from the genome to the sequence in ZoomMap.

Figure 2. Use of magic lenses in ZoomMap.

ZoomMap

In ZoomMap (11), HuGeMap is represented as a two-dimensional world where the user can travel by zooming in and out on objects. The starting point is a rough karyotype of the human genome (Fig. 1a). As the user zooms in, more details appear as chromosomes get closer and space becomes available. The arms of the chromosomes are visible on Figure 1b, later the cytogenetic banding and textual information (names of the maps of the chromosome) are displayed (Fig. 1c). When zooming in further,
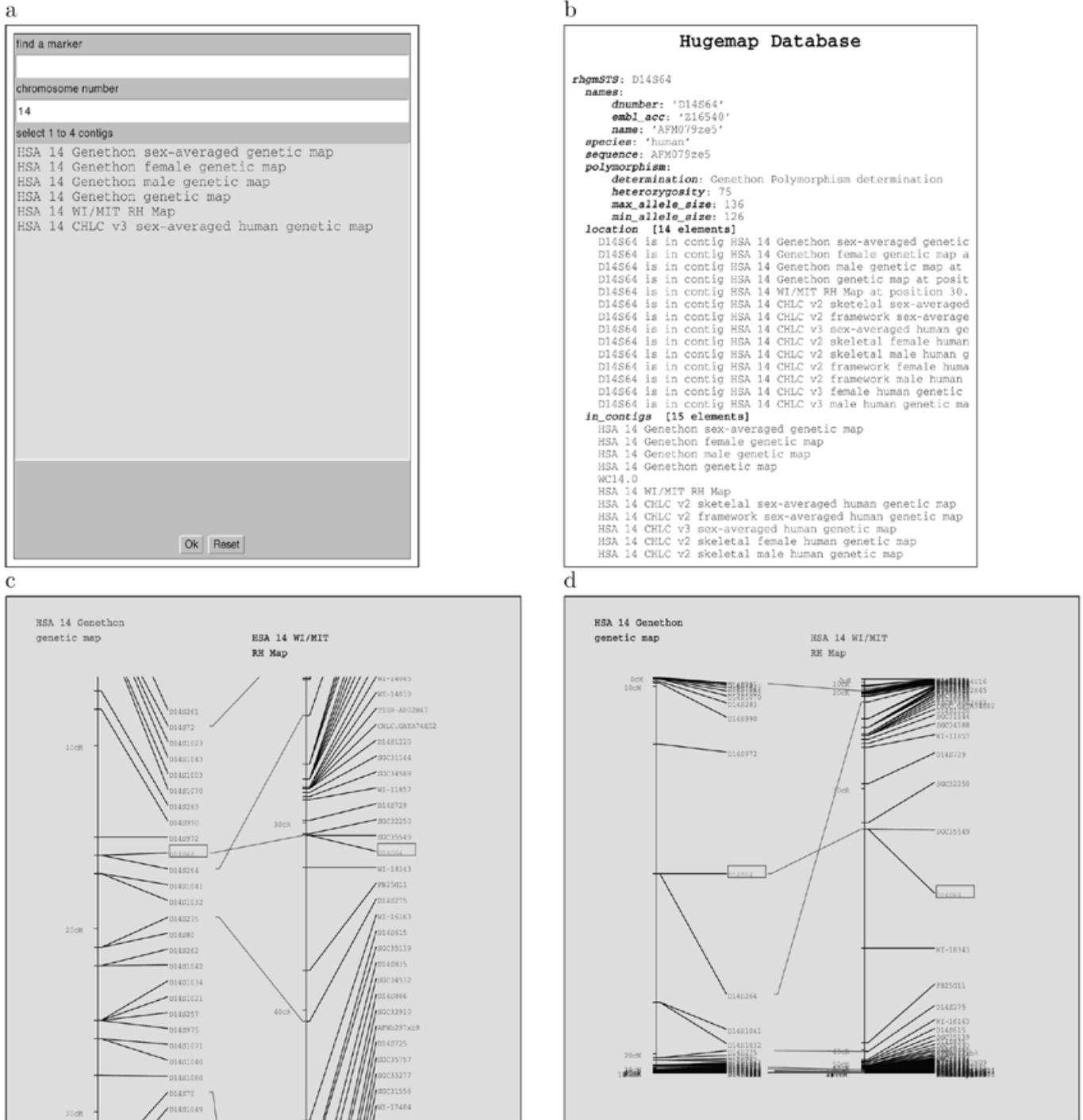
Figure 3. Visualizing maps with MappetShow.

the name of the chromosome maps are replaced by the usual representation of a map (Fig. 1d) with more and more markers as scale increases (Fig. 1e). Finally the sequence appears (Fig. 1f).

In Figure 1e, the square boxes linked to a marker are portals which display another part of the HuGeMap world: they point to the other maps on which the marker is positioned. They allow the user to navigate directly to a related map.

The data represention used by ZoomMap can be modified by the use of magic lenses. These are dynamic items under user control that allow a transformation of the data display in the box defined by the lens. Figure 2a shows two lenses used to convert the chromosome banding into chromosome marker maps. In Figure 2b, the bottom lens displays the marker names using a grey scale depending on their polymorphism and the top lens gives

more information on the markers. In Figure 2c, a lens that filters out the markers with low polymorphism has been stacked on the bottom lens.

ZoomMap was written with the Zomit package. It can be used with the Zomit Java applet client at http://www.infobiogen.fr/ services/zomit/ZoomMap.html

MappetShow

MappetShow (Fig. 3) is a graphical interface, written in Java, that draws and compares maps (http://www.infobiogen.fr/services/ Mappet). The user can choose a chromosome or a marker. The maps related to the chromosome or containing the marker are listed and can be selected for display (Fig. 3a). Each displayed map can be browsed and scaled independently (Fig. 3c). When the user specifies a marker, all maps are recentred on this marker. MappetShow is also connected to the HuGeMap WWW browser (Fig. 3b).

In MappetShow we implemented a Focus+Context technique which shows a detailed view of the area of interest, while keeping a summary of the remaining data visible. This is performed with variable scaling. It allows the user to focus on the markers of interest of a map, while keeping a view of the context of the whole map.

Independent non-linear views facilitate browsing and graphical comparison of different maps on different regions (Fig. 3d).

PERSPECTIVES

Our intention is to integrate all mammalian genome maps with the human maps. MappetShow and ZoomMap will be modified to display the comparative maps. We will use our standard genome map IDL in the development of a mammalian genome database at Infobiogen and promote the use of this standard format in the mapping community under the auspices of the Object Management Group (http://www.omg.org).

ACKNOWLEDGEMENTS

REFERENCES

1 Dib,C., Fauré,S., Fizames,C., Samson,D., Drouot,N., Vignal,A., Millasseau,P., Marc,S., Hazan,J., Seboun,E. *et al.* (1996) *Nature*, **380**, 152–154.
2 Sheffield,V.C., Weber,J.L., Buetow,K.H., Murray,J.C., Even,D.A., Wiles,K., Gastier,J.M., Pulido,J.C., Yandava,C., Sunden,S.L. *et al.* (1995) *Hum. Mol. Genet.*, **4**, 1837–1844.
3 Bellanné-Chantelot,C., Lacroix,B., Ougen,P., Billaut,S., Beaufils,S., Bertrand,S., Georges,I., Glibert,F., Gros,I., Lucotte,L. *et al.* (1992) *Cell*, **70**, 1059–1068.
4 Cohen,D., Chumakov,I. and Weissenbach,J. (1993) *Nature*, **336**, 698–701.
5 Chumakov,I., Rigault,P., Le Gall,I., Bellanné-Chantelot,C., Billault,A., Guillou,S., Soularue,P., Guasconi,G., Poullier,E., Gros,I. *et al.* (1995) *Nature*, **377**, 175–298.
6 Hudson,T.J. *et al.* (1995) *Science*, **270**, 1945–1954.
7 Harnden,D.G. and Klinger,H.P. (eds) (1985) *An International System for Cytogenetic Nomenclature.* S. Karger AG, New York.
8 Lijnzaad,P., Helgesen,C. and Rodriguez-Tomé,P. (1998) *Nucleic Acids Res.*, **26**, 102–105.
9 Achard,F. and Barillot,E. (1997) In Altman,R., Dunker,K., Hunter,L. and Klein,T. (eds), *Pacific Symposium on Biocomputing ′97.* World Scientific, pp. 39–50.
10 Barillot,E., Leser,U., Lijnzaad,P., Cussat-Blanc,C., Jungfer,K., Guyon,F., Vaysseix,G., Helgesen,C. and Rodriguez-Tomé,P. (1998) *Bioinformatics*, in press.
11 Pook,S., Vaysseix,G. and Barillot,E. (1998) *Bioinformatics*, **14**, 807–814.