A set of ontologies to drive tools for the control of vector-borne diseases

Pantelis Topalis[1], Emmanuel Dialynas[1], Elvira Mitraka[2], Elena Deliyanni[1], Inga Siden-Kiamos[1] and Christos Louis[1,2]

[1]Institute of Molecular Biology and Biotechnology, Foundation for Research and Technology-Hellas, 711 10 Heraklion, Crete, Greece; [2] Department of Biology, University of Crete, 711 10 Heraklion, Crete, Greece

Abstract

We are developing a set of ontologies that deal with vector-borne diseases and the arthropod vectors that transmit them. For practical reasons (application priorities), we initiated this project with an ontology of insecticide resistance followed by a series of ontologies that describe malaria as well as physiological processes of mosquitoes that are relevant to, and involved in, disease transmission. These will be expanded to encompass other vector-borne diseases as well as non-mosquito vectors. The aim of the whole undertaking, which is worked out in the frame of the international IDO (Infectious Disease Ontology) project, is to provide the community with a set of ontological tools that can be used both in the development of specific databases and, most importantly, in the construction of decision support systems to control these diseases.

The problem of vector-borne diseases

Epidemiologists have brought together in one "functional "group a series of diseases of different etiology and pathogenesis that share one key component: their mode of transmission (see Goddard, 1999[1] and several chapters of Marquardt, 2005[2] for specific questions addressing insect-borne diseases and their vectors). These diseases are transmitted by the bite of a specific arthropod vector, which is usually an insect. The pathogenic agent is passed with the saliva transferred during the bite to the potential patient. Two additional characteristics are shared by most vector-borne diseases, namely most people affected live in the tropical regions of the world and, connected to this, the diseases affect mostly populations that are also heavily affected by poverty. The pathogens responsible for these diseases are very diverse, ranging from protozoan parasites (e.g. *Plasmodium* spp. in malaria, *Leishmania* spp. in leishmaniosis) and bacteria (e.g. *Borrelia* spp. in Lyme disease) to worms (e.g. Nematodes in filariasis and river blindness) to viruses (e.g. Dengue, Yellow fever). Similarly, the vectors range from mosquitoes (e.g. malaria and Dengue) and flies (e.g. Tsetse in African trypanosomiasis) to kissing bugs (Chagas' disease) and ticks (e.g. Lyme disease). The extreme variation in the biology of both pathogens and the vectors makes it difficult to address vector-borne diseases as a whole. Importantly, these difficulties also affect important aspects such as prevention, epidemiology, therapy, etc.

A common theme, which in a sense unites these diseases, is the fact that their transmission can be blocked if the agents that transmit them, i.e. the arthropod vectors, are removed from the pertinent chain of events[3]. Vector control has therefore historically become a *conditio sine qua non* for the control of these infections[4,5], and this fact has been exemplified by the elimination of malaria from most non-tropical areas of the globe[6]. While leading to about half a billion cases in the tropics every year, and still being responsible for anything between one and three million deaths (mostly children in sub-Saharan Africa), this killer illness has practically disappeared from Europe and North America through intense insecticidal measures aimed at eliminating the *Anopheles* vectors. It should be stressed that, with the exception of the Yellow fever[7], no vaccine is currently available for any vector-borne disease as an alternative prevention strategy that would act on a different level than that of the actual vector. Prevention focused on the vector includes not only control of insect populations through environmental management or the use of chemicals, but also the protection of individuals through the use of clothing, repellents, nets and screens[8].

Although greatly successful in the previous century, insect-control programmes are now immensely obstructed by a variety of reasons. These range from community opposition to a vast usage of chemicals[9], to the development of resistance against these very chemicals by the insect vectors to be controlled[10]. Moreover, these problems are aggravated by several facts: resistance against drugs is also encountered in the pathogens[11]; vaccine development, if at all possible, is slow[12]; new drug development is not only slow but extremely expensive and the areas affected by the diseases in question are certainly not the ones that can easily spearhead such efforts due to the lack of economic and scientific resources in these areas[13]. It is therefore of utmost importance to develop

innovative strategies for the control of vector-borne diseases. One novel approach is to use IT technologies as a complement to the application of developments in the biology of disease vectors. While the latter projects make use of scientific research products such as whole genome sequences[14,15], transgenesis[16], and the use of other "intelligent" approaches[17] the former potentially brings new specific tools that can be used for a more efficient, and often close-to-the-field management of pertinent disease data, especially entomological ones.

In this context, our group has embarked on a long project that involves the development of ontologies dealing with disease vectors and vector-born diseases[18,19]. The obvious rationale behind is the need of these ontologies to unify the "language" spoken by vector biologists and epidemiologists. The ultimate end is to build a comprehensive ontology for insect-borne diseases that may consist of sub-ontologies, each addressing a specific aspect of the whole. In the frame of the Infectious Disease Ontology project (http://www.infectiousdiseaseontology.org/Home.html), we initiated this effort focusing on malaria, but we are already expanding this to encompass the other diseases of this class as well. These ontologies, some of which are already available and some under construction, will be presented below in a summary form.

Ontologies and vector-borne diseases: a brief description

The aspects of vector-borne diseases that are in need of an ontological description range from those that deal with the diseases as such (e.g. pathogenesis, clinical aspects, therapy, etc.), to vector biology (physiological processes of the vectors) and to epidemiology and control in the widest sense of the terms (prevention, insect control, etc.). As stated earlier, these aspects are extremely diverse and complex, simply given the multitude of organisms involved (vectors and pathogens in addition to the human host) and the fact that we are often dealing with populations (additional level of granularity!). The construction of a comprehensive ontology, thus, if at all feasible, must be addressed using a piecemeal approach. It is clear that certain fundamental decisions have to be taken at the initial phases, and an open-ended advance is, in our mind, a must. In that sense, we decided, early on, that the end product would have to follow i) the rules set by the OBO Foundry[20] and ii) be based on the basic formal ontology[21,22]. If long-term interoperability of future databases is to be achieved, these two choices are a prerequisite. This rule, of course, is the end goal and we decided to keep a certain degree of flexibility throughout the project until a "unified" ontology is constucted. One example for such a flexible approach is the fact that the ontology of insecticide resistance in mosquitoes that we developed (MIRO) does not follow the BFO in its initial versions but, rather is structured such that it can be adopted without many problems by the community that immediately needs to apply it in the field (Topalis et al., 2009, submitted). The MIRO forms the core of the related database on insecticide resistance (IRbase) that we also developed, and which was adopted for immediate use by the World Health Organization (Topalis et al., 2009, submitted). We should state that we are nevertheless in the process of long term restructuring the ontology along BFO standards, such that its contents can be later included in the comprehensive ontology on vector-borne diseases.

Although already submitted to and listed by the OBO Foundry (http://www.obofoundry.org/cgi-bin/detail.cgi?id=mosquito_insecticide_resistance), MIRO is a pure application ontology that is being used to drive a dedicated database, IRbase (http://anobase.vectorbase.org/ir/). It consists of four specially devised sub-ontologies that cover all aspects of inscticide resistance, with an emphasis on field work and monitoring. Thus, although mechanisms of resistance are covered, this is not done in detail. Furthermore, MIRO's fifth component, a geographical one, uses *in toto* the controlled vocabulary Gazeteer (http://darwin.nerc-oxford.ac.uk/gc_wiki/index.php/GAZ_Project) to provide IRbase curators with records describing the areas in which data were collected. The MIRO is constantly being updated upon request by members of the international community that is involved in the study of insecticide resistance.

The second ontology, which is still nameless, covers physiological processes of mosquitoes that are involved in disease transmission. The processes covered do not only address the actual transmission, i.e. the interplay between vectors and pathogens but, importantly, also the actual progression of events in the vector. We want to stress that the processes mentioned here are, in their vast majority, processes on the level of the organism and not cellular or sub-cellular ones, such as the ones covered by the GO[23, 24]. Thus, (near) top level classes are, among others, behaviour, sensory perception, processes of the immune system and nutrition. As an example, when looking at the children of "behaviour", one will find a line of terms leading, through the adult feeding behaviour to entities such as the four phases of "interrupted feeding" (exploratory phase, imbibing phase, probing phase and withdrawal phase). The ontology also covers processes that are not directly

"linked" to disease transmission and this, obviously, for reasons of completion. For reasons of orthogonality, in all cases in which terms are already covered by established ontologies, we adhere to these, along with their descendants. This is notably the case for the Processes sub-ontology of the GO. Our ontology is far from complete, although it already covers more than 600 terms, which are all fully defined.

The next ontology that we are in the process of populating with terms is the one describing malaria. This is the actual ontology that we decided to develop in the frame of IDO, and which we plan to expand in the near future in order to cover other vector-borne diseases as well. It is built based on BFO and the IDO reference ontology (http://www.infectiousdiseaseontology.org/IDO_files/IDO_10.08.07.obo.txt), and it is meant to cover malaria on all possible levels. These obviously include both the clinical aspects of the disease in the widest sense (i.e. including epidemiology, etc.) and the biology of the disease that describes processes and objects of not immediate clinical relevance. We consider as such items like proteins involved in the penetration of both mosquito and human/vertebrate cells as well as their interacting partners in the *Plasmodium* parasites. Again, similarly to the case of the ontology of physiological processes, we have taken care to include wherever possible direct imports of pre-existing ontologies. One such example is the Plasmodium parasite life cycle stage and its descendants that all have crossreferences to the, at the moment, inactive Plasmodium life cycle ontology. The malaria ontology has at this time about 600 terms.

Ontologies and vector-borne diseases: concluding remarks

The ontologies that we are constructing can be described as pure application ontologies that are meant to form the basis for specific tools such as specific databases or decision support systems for various diseases. The need for such tools became apparent immediately after the first working version of the MIRO and its "cousin" IRbase were made public. Not only did the international community immediately decide to adopt both tools, but already within a few months after the initiation of data population, there are about 800 sampled populations that are shown in the database. This is about 683 more than what the insecticide resistance section in VectorBase carried, the only repository for data of this kind. In addition to databases that are driven by ontologies in an increasing fashion (see for example databases using the ontology-depending schema Chado[25], such as FlyBase[26,27] and VectorBase[28,29], ontologies are ideal tools for the design of intelligent decision support systems. In cases such as vector-borne diseases, whose control is also hampered by weak infrastructure in endemic countries, these DSSs can be used by medical workers and health agencies in remote areas, either for ongoing studies or in cases that need immediate attention[30, 31] .

One of the intricacies that we are already faced with is the planned expansion of the malaria-oriented ontologies, to cover many other vector-borne diseases. To understand the magnitude of the problem one should think of the fact that vector-borne diseases represent major threats to public health in wide and ecologically diverse areas of the world, are caused by completely different pathogens and are transmitted by completely different vectors. Thus, the challenge now is how to cover this broad spectrum of facts in a single ontology. There is naturally the possibility of cutting through the Gordian knot, by devising separate ontologies for each disease. The counter-argument in this case would be that, brought to an extreme, each malaria form (i.e. tertian, malignant and benign, and quartan, should have its own ontology) similar to the different forms of filariasis that are caused by different species of nematodes and whose clinical aspect differ only slightly. In addition, similarities between these diseases and the agents that transmit them may be obscured if different ontologies were used and this would certainly have a negative impact on their value in the long term. Therefore, we are still trying to solve the knot in a non-Alexandrian way. By "merging" the ontologies into one, we can also actively support the rules of the OBO Foundry and provide an example of how the construction of a large and comprehensive ontology can, later on, provide advantages to its users.

Acknowledgements

The work was funded by contract HHSN266200400039C from the National Institute of Allergy and Infectious Diseases in the frame of the VectorBase project and by the BioMalPar European Network of Excellence supported by a European grant (LSHP-CT-2004-503578) from the Priority 1 "Life Sciences Genomics and Biotechnology for Health" in the 6th Framework Programme. The authors would numerous colleagues who helped at different stages of the work, and in particular Drs. John Vontas for his contribution to the MIRO, Frank Collins for his encouragement and support in the frame of VectorBase and Barry Smith and Lindsay Cowell for accepting us in the IDO community.

References

1. Goddard J. Infectious diseases and arthropods. Totowa, N.J.: Humana Press; 2000.
2. Marquardt WC, Kondratieff BC. Biology of disease vectors. 2nd ed. Burlington, MA: Elsevier Academic Press; 2005.
3. Hemingway J, Beaty BJ, et al. The Innovative Vector Control Consortium: improved control of mosquito-borne diseases. Trends Parasitol 2006;22: 308-12.
4. della Torre A, Arca B, et al. The role of research in molecular entomology in the fight against malaria vectors. Parassitologia 2008;50: 137-40.
5. Peter RJ, Van den Bossche P, et al. Tick, fly, and mosquito control--lessons from the past, solutions for the future. Vet Parasitol 2005;132: 205-15.
6. de Zulueta J. The end of malaria in Europe: an eradication of the disease by control measures. Parassitologia 1998;40: 245-6.
7. Roukens AH, Visser LG. Yellow fever vaccine: past, present and future. Expert Opin Biol Ther 2008;8: 1787-95.
8. Hill J, Lines J, Rowland M. Insecticide-treated nets. Adv Parasitol 2006;61: 77-128.
9. Schapira A. DDT: a polluted debate in malaria control. Lancet 2006;368: 2111-3.
10. Hemingway J, Ranson H. Insecticide resistance in insect vectors of human disease. Annu Rev Entomol 2000;45: 371-91.
11. Laufer MK. Monitoring antimalarial drug efficacy: current challenges. Curr Infect Dis Rep 2009;11: 59-65.
12. Langhorne J, Ndungu FM, et al. Immunity to malaria: more questions than answers. Nat Immunol 2008;9: 725-32.
13. Craft JC. Challenges facing drug development for malaria. Curr Opin Microbiol 2008;11: 428-33.
14. Holt RA, Subramanian GM, et al. The genome sequence of the malaria mosquito Anopheles gambiae. Science 2002;298: 129-49.
15. Nene V, Wortman JR, et al. Genome sequence of Aedes aegypti, a major arbovirus vector. Science 2007;316: 1718-23.
16. James AA. Preventing the spread of malaria and dengue fever using genetically modified mosquitoes. J Vis Exp 2007: 231.
17. Rasgon JL. Using predictive models to optimize Wolbachia-based strategies for vector-borne disease control.Adv Exp Med Biol 2008;627: 114-25.
18. Topalis P, Tzavlaki C, et al. Anatomical ontologies of mosquitoes and ticks, and their web browsers in VectorBase. Insect Mol Biol 2008;17: 87-9.
19. Topalis P, Lawson D, Collins FH, Louis C. How can ontologies help vector biology? Trends Parasitol 2008;24: 249-52.
20. Smith B, Ashburner M, et al. The OBO Foundry: coordinated evolution of ontologies to support biomedical data integration. Nat Biotechnol l2007;25: 251-5.
21. Simon J, Dos Santos M, Fielding J, Smith B. Formal ontology for natural language processing and the integration of biomedical databases. Int J Med Inform 2006;75: 224-31.
22. Grenon P, Smith B, Goldberg L. Biodynamic ontology: applying BFO in the biomedical domain. Stud Health Technol Inform 2004;102: 20-38.
23. Ashburner M, Lewis S. On ontologies for biologists: the Gene Ontology--untangling the web. Novartis Found Symp 2002;247: 66-80; discussion 80-3, 84-90, 244-52.
24. Harris MA, Clark J, et al. The Gene Ontology (GO) database and informatics resource. Nucleic Acids Res 2004;32: D258-61.
25. Mungall CJ, Emmert DB. A Chado case study: an ontology-based modular schema for representing genome-associated biological information. Bioinformatics 2007;23: i337-46.
26. Gelbart WM, Crosby M, et al. FlyBase: a Drosophila database. The FlyBase consortium. Nucleic Acids Res 1997;25: 63-6.
27. Tweedie S, Ashburner M, et al. FlyBase: enhancing Drosophila Gene Ontology annotations. Nucleic Acids Res 2009;37: D555-9.
28. Megy K, Hammond M, et al. Genomic resources for invertebrate vectors of human pathogens, and the role of VectorBase. Infect Genet Evol 2008.
29. Lawson D, Arensburger P, et al. VectorBase: a data resource for invertebrate vector genomics. Nucleic Acids Res 2009;37: D583-7.
30. Thomson MC, Connor SJ, et al. The ecology of malaria--as seen from Earth-observation satellites. Ann Trop Med Parasitol 1996;90: 243-64.
31. Coleman M, Sharp B, et al. Developing an evidence-based decision support system for rational insecticide choice in the control of African malaria vectors. J Med Entomol 2006;43: 663-8.