

# Bioinformatics: Current perspectives and future directions for food and nutritional research facilitated by a Food- Wiki database

Thérèse A. Holton<sup>a,b,1</sup>,  
Vaishnavi Vijayakumar<sup>a,b,1</sup> and  
Nora Khaldi<sup>a,c,\*</sup>

<sup>a</sup>UCD Complex and Adaptive Systems Laboratory/  
UCD Conway Institute of Biomolecular and  
Biomedical Science/School of Medicine and Medical  
Science, University College Dublin, Dublin 4, Ireland

<sup>b</sup>Food for Health Ireland, University College Dublin,  
Dublin 4, Ireland

<sup>c</sup>Food Science and Technology, University College  
Davis, CA, USA (UCD Complex and Adaptive Systems  
Laboratory, University College Dublin,  
Dublin 4, Ireland. Tel.: +353 1 7165335;  
e-mails: [nora.khaldi@ucd.ie](mailto:nora.khaldi@ucd.ie); [khaldin@tcd.ie](mailto:khaldin@tcd.ie))

With the continued progression of the “omics” era, bioinformatics, a discipline concerned with the curation and interpretation of biological data by computational means, has seen widespread integration across life sciences. However, despite becoming very data rich disciplines in the past number of years, the role of bioinformatics in food and nutritional sciences is less appreciated. In this review, we present the current state of bioinformatics in food and nutritional sciences and offer exemplar templates from other fields where bioinformatic

analyses have become an integral feature. Additionally, we propose the concept of a wiki-like food database that could greatly advance the capabilities of bioinformatics in food and nutritional research.

## Introduction

Biology is a knowledge based discipline, now more so than ever, with various integral techniques producing data at an inexorable rate. For example, as of June 2013, GenBank, which is the main genetic sequence repository, has approximately 164 million gene sequence records and approximately 110 million genome sequence entries (<http://www.ncbi.nlm.nih.gov/genbank/statistics>). Further to this, Uniprot Knowledgebase (Uniprot), which is the primary protein sequence and function resource, has over 13.5 million entries as of January 2011 (Magrane & Consortium, 2011). With the accumulation of such volumes of data, the use of computers has become imperative for its curation and interpretation. This has given rise to new field of research, namely bioinformatics, where computer science and statistics are applied to analyse biological data with the aim of expediting, ameliorating, and diversifying biological research.

Bioinformatics has been established as an important scientific discipline, bringing about a paradigm shift in various disciplines including molecular medicine, comparative genomics, molecular evolution, microbial genome applications, drug discovery and biotechnology. However, its applicability in the food science arena is less appreciated, despite the recognised potential for a significant contribution to this field (Desiere, German, Watzke, Pfeifer, & Saguy, 2001). In this review, we highlight instances where bioinformatics has been successfully applied in food and nutrition sciences to advance and expand research. Various factors, such as allergenicity, flavour, functionality, probiotics and prebiotics, influencing both food scientists and industry alike, are outlined. Additionally, we will discuss how the use of bioinformatics in food is still in its infancy, and therefore propose the concept of a Food-Wiki database (FoodWikiDB). In this new database, food and nutritional information will be stored in a consensus style, making the utilisation of this immense amount of data possible and easily managed by bioinformatic strategies and protocols. We feel such a resource will advance and develop food and nutritional sciences, with a view to improving the quality and nutritive value of food sources.

\* Corresponding author.

<sup>1</sup> Joint first author.

## Bioinformatics – general applications and implementations

Bioinformatics can essentially be defined as the application of computational resources to biological data. It is largely concerned with data types such as DNA and protein sequences, as well as protein domains and structures. Owing to this very general definition and diverse data spectrum, it follows that bioinformatics has an equally broad and varied remit within life sciences.

There have been many success stories attributed to the use of bioinformatics approaches and here we outline but two of the most notable areas, specifically comparative genomics and drug discovery. Learnings from these areas have the potential to be valuable in a food and nutritional science setting.

### Comparative genomics

Considerable advancements have, and continue to be made in genome sequencing technologies. Accordingly, this, coupled with marked improvement in compute power, has contributed greatly to advancements in comparative genomics. As the name suggests, comparative genomics is simply the analysis of similarities and variations between genomes, meaning the applications of this field are unsurprisingly extremely diverse. Such approaches can reveal information about the structure and size of genomes across a domain of life (Bentley & Parkhill, 2004; Rubin *et al.*, 2000), elucidation of important genetic elements (Khaldi & Wolfe, 2008; Khaldi & Wolfe, 2011; Ma *et al.*, 2010; Wassarman, Repoila, Rosenow, Storz, & Gottesman, 2001), modes of evolution (Ogura *et al.*, 2009), genetic components involved in capacities such as pathogenicity (Carlton *et al.*, 2008; Chun *et al.*, 2009; Fedorova *et al.*, 2008; Khaldi, Collemare, Lebrun, & Wolfe, 2008; Khaldi *et al.*, 2010; Khaldi and Wolfe, 2008) and environmental adaptations (Omelchenko *et al.*, 2005).

In a more practical context, comparative genomics has been applied to attenuated strains of bacteria in existing vaccines in an attempt to develop improved vaccines (Behr *et al.*, 1999). Additionally, this field has helped ascertain the genes involved in biofuel production in fungi (Wohlbach *et al.*, 2011), while in plants the exploration of their potential in this regard is ongoing (Paterson *et al.*, 2009; Rubin, 2008). Many food pathogens, and indeed food and food-related sources, have been the focus of sequencing, and consequently comparative genomic projects (Choi *et al.*, 2004; Doumith *et al.*, 2004; Makarova *et al.*, 2006; Mayer *et al.*, 2011; Paterson, Bowers, & Chapman, 2004). It is perceived that such studies may contribute to the prevention of crop diseases (Qian *et al.*, 2005) and food poisoning (Abee, Van Schaik, & Siezen, 2004), in addition to augmenting the knowledge of important processes and properties such as ripening (Fei *et al.*, 2004), flavour formation (Liu, Nauta, Francke, & Siezen, 2008) and crop improvement (Mochida & Shinozaki, 2010; Varshney, Graner, & Sorrells, 2005).

### Drug discovery

In the last number of years, bioinformatics has been established as a major player in the drug discovery process (Buchan *et al.*, 2011; Butte & Ito, 2012; Taboureau, Baell, Fernández-Recio, & Villoutreix, 2012). It is now being utilised as standard practice in both academic and industrial settings, significantly reducing the duration of many stages of the lengthy drug discovery process. With the wealth of sequencing data accrued from the genomic era, there is great potential for detecting novel drug targets. Indeed, it is estimated that there is approximately 3000 potential drug targets encoded in the human genome alone (Hajduk, Huth, & Tse, 2005; Hopkins & Groom, 2002). Accordingly, bioinformatic approaches have been developed to determine targets within “the druggable genome”, as well as in additional sources such as biomarkers (Colburn, 2003), pathways (Demir *et al.*, 2002) and genetic elements such as miRNA (Mazière & Enright, 2007). Structural bioinformatics has also been successful for the identification of drug targets based on protein structure homology and comparative modelling procedures (Blundell *et al.*, 2006).

Similarly, bioinformatics has been used for the identification of lead compounds in drug discovery. For example, bioinformatic techniques are used to improve high throughput screening (HTS) techniques of thousands of compounds (Makarenkov *et al.*, 2007). Further to this, structural bioinformatics approaches such as docking methods (Taylor, Jewsbury, & Essex, 2002) and pharmacophore searching are proving to be both companion, and potentially alternative, tools to high throughput screening (McInnes, 2007). The scope of bioinformatics in both target and lead identification may well be attributed to the numerous centralised resources (Berman *et al.*, 2000; Gaulton *et al.*, 2012; Wishart *et al.*, 2006) at the disposal of the community. The successful application of bioinformatics in the area of drug discovery is particularly pertinent to food and nutritional industry, demonstrating very well the transition from an *in silico* to applied setting.

### Why bioinformatics for food?

Food plays a vital role in maintaining wellbeing by regulating metabolic, hormonal, physical and mental processes. Moreover, there is an ever-increasing appreciation of the role played by nutrition in the progression of various chronic diseases. Great efforts are now being focused to promote and augment the quality and nutritive potential of various food sources. The concept of food as means of preventing or alleviating medical conditions has evidently translated to the spending habits of consumers, with the global nutraceutical market expected to reach US\$243 billion by 2015 (Global Industry Analysts Inc., 2012). In line with this, food science has evolved significantly, embracing many emerging techniques such as the “omics” suite of technologies for example. With such a movement towards diversification, this discipline is now facing the

challenge of interpreting and integrating many varied data types. In the subsequent section of this review we describe a number of ways in which bioinformatics is being incorporated into food and nutrition sciences.

Although there has been a definite and concerted adoption of bioinformatics in these fields, there are a number of issues outstanding as the shift towards computation continues. Firstly, we feel there is a need for education on the potential and capabilities, as well as an awareness of the limitations, of the applications of bioinformatics in food and nutrition. Secondly, we suggest the introduction of a common and unified language to provide researchers with disparate backgrounds a common platform to communicate from. Finally, we feel it is necessary to establish an integrative database to store and curate the wide range of biological data and resources that are important to food and nutritional sciences. We, like others (Lemay, Zivkovic, & German, 2007), consider these measures vital to the current progression of bioinformatics in food as they will provide for a simple and expedient way for researchers to embrace a new discipline. A more detailed outline of these proposals can be found in the ensuing sections.

At this point, bioinformatics is poised to significantly help the growing focus of academics, industry and consumers on improving the quality and functionality of food sources. This field can ameliorate many established procedures and protocols in food and nutrition sciences, as well as provide a means for the development of novel and innovative interdisciplinary techniques.

### Food bioinformatics: the story to date

Despite the limited availability of integrative databases and dedicated resources, many bioinformatic approaches have, and continue to be established in food and nutrition sciences. In the sections below, we outline some of the ways bioinformatics has been successfully integrated into these disciplines.

#### 'Omics' technologies

As is the case in many other biological domains, the need for the development of specific bioinformatic strategies has largely stemmed from the advent of high throughput "omics" techniques. Such "omics" technologies, which include genomics, proteomics, metabolomics, and more recently metagenomics, have contributed significantly to the amassment of biological data. Suitably, many have turned to bioinformatic approaches to manage and interpret the volumes of data accrued from these so-called "foodomics" technologies (Cifuentes, 2009; Herrero, García-Cañas, Simo, & Cifuentes, 2010).

Proteomics and bioinformatics have become intrinsically linked fields of research, and so it is not surprising that they are emerging as strong partners in the foodomics era also. Proteomics, in a food and nutrition context, is fast becoming an invaluable tool to allow for the study of protein concentrations of important foods, as well as their

associated post-translational modifications (Amiour, Merlino, Leroy, & Branlard, 2002). By adopting a combined proteomic and bioinformatic framework it has been possible, for example, to elucidate more comprehensive protein compositions (including functional information) of food sources (D'Alessandro, Zolla, & Scaloni, 2011), and to develop novel approaches for the characterisation of foodborne micro organisms (Fagerquist *et al.*, 2010; Fagerquist *et al.*, 2009).

In the field of metabolomics, which studies the entire (or at least a substantial amount of the) metabolite profile of a cell, bioinformatics is becoming an equally valuable tool. Unsurprisingly, metabolomics has the potential to produce considerable data sets, particularly in its "un-targeted" form (Orešič, 2009). Bioinformatic tools, designed for the management, curation, analysis and visualization of metabolic data include Mzmine (Katajamaa, Miettinen, & Orešič, 2006), MetAlign (Lommen, 2009) and MetExtract (Bueschl *et al.*, 2012). In addition to software development, bioinformatics is being applied in a more direct fashion, for example, in the discovery of metabolites associated with diabetes (Altmaier *et al.*, 2008). Recently, bioinformatics has additionally aided in the integration of metabolomics into a genome wide association study (GWAS) context, where metabolite ratios are used in combination with single nucleotide polymorphisms (SNPs) to determine genetic variants that are related to certain metabolic traits (Gieger *et al.*, 2008; Illig *et al.*, 2010).

#### Bioactive peptides

It is now known that various food-derived peptides exhibit a range of functional activities that exceed their fundamental nutritional role. Consequently, there has been increased academic and commercial interest in the use of foods enriched in bioactive peptides for the promotion of health and alleviation of various conditions (Hartmann & Meisel, 2007; Khaldi, 2012; Korhonen & Pihlanto, 2006). For the most part, bioactive peptides of food origin are released from their precursor proteins either by bacterial fermentation or through the action of gastrointestinal or other proteolytic enzymes. While many methods have been developed to detect and characterise peptides produced in this manner (Gibbs, Zougman, Masse, & Mulligan, 2004; Matsui, Li, & Osajima, 1999; Mullally, Meisel, & Fitzgerald, 1997; Vermeirssen, Van Camp, & Verstraete, 2002), it has become apparent that the use of bioinformatics can serve as a valuable strategy in the discovery of novel bioactive peptides (for review see Khaldi, 2012).

There have been many approaches developed for the prediction of bioactive peptides (see Fig. 1) which, although designed for a more general scope, have shown some success in food sources. Examples include computational methods devised for the prediction of antimicrobial peptides (Torrent, Andreu, Nogués, & Boix, 2011; Torrent, Nogués, & Boix, 2009). Others have used a sequence homology based approach to identify peptides with various properties in chicken (Lynn *et al.*, 2004;



As many foodborne pathogens have been the focus of genomic sequencing projects, there is a growing appreciation for the potential of bioinformatics in the area of food safety and quality (Brul *et al.*, 2006). For example, the Food and Drug Administration (FDA) have recently developed a bioinformatics based tool for detecting and identifying bacterial food pathogens (Fang *et al.*, 2011). Further to this, the onset of next generation sequencing technologies has provided for a novel way to bioinformatically determine the source of outbreaks of foodborne illness (Lienau *et al.*, 2011). Other computational applications in this area have included the use of neural networks with the aim of predicting microbial growth within a given food source (García-Gimeno, Hervás-Martínez, & De Silónizc, 2002). With regard to food quality, great progress continues to be made in the area of computer vision which allows for the automated appraisal of various food properties (Mery, Pedreschi, & Soto, 2012).

#### Allergen detection

Since allergens from very diverse and disparate sources have high sequence and structural similarity, often causing comparable IgE responses, it is not surprising that bioinformatics poses an increasingly important resource in the study of food allergens. Many dedicated allergen databases with a specific focus on food exist, for example AllerMatch (Fiers *et al.*, 2004), Informall (<http://foodallergens.ifr.ac.uk>), FARRP Allergen database (<http://allergenonline.org>) and SDAP (Ivanciuc, Schein, & Braun, 2003; Mari *et al.*, 2006), which has greatly facilitated the applicability of bioinformatics in this area.

Common bioinformatic homology searches, in addition to structural bioinformatics approaches, are routinely executed to establish the potential allergenicity and cross reactivity of proteins (Baderschneider *et al.*, 2002; Hileman *et al.*, 2002; Jenkins, Griffiths-Jones, Shewry, Breiteneder, & Mills, 2005). Indeed, the utility of such approaches has prompted the World Health Organisation (WHO) to include sequence similarity searches as part of the guidelines for assessing allergenicity of genetically modified foods (FAO/WHO, 2001). Other bioinformatic research interest in this area includes the development of tools to predict allergenicity using support vector machines (Saha & Raghava, 2006), common sequence based motifs (Ivanciuc, Garcia, Torres, Schein, & Braun, 2009) and amino acid properties (Dimitrov, Flower, & Doytchinova, 2013). Differently, bioinformatics techniques have recently been implemented in the development of allergen diagnostics where machine learning was applied to the prediction of peanut allergy (Lin *et al.*, 2012).

#### Pro and prebiotics

Greater insight into the diverse microbial flora present in the human gastrointestinal tract (GIT), which are responsible for healthy intestinal function and resistance to numerous diseases (Fuller, 1989), has led to considerable

marketing of pro and prebiotic foods. In a similar manner to many of the topics discussed above, the advent of the genomic era has greatly propelled the use of bioinformatics in the mechanistic study of pro and prebiotics. For example, genome sequencing of probiotic intestinal bacteria has helped identify genes potentially involved in GIT interactions and highlight various adaptations to the GIT environment (Pridmore *et al.*, 2004). Additionally, genome analysis of various strains of Bifidobacteria has revealed specific evolutionary adaptations to the human oral cavity (Ventura *et al.*, 2009) and distal gut (Turroni *et al.*, 2009). Combined with mass spectroscopy and molecular technologies, recent bioinformatic applications have enabled the study of gut microbiota providing insights into metabolites present due to interaction with the host (Rath *et al.*, 2012). Through the advances in high throughput sequencing in the area of probiotics, some researchers can now propose the selection of probiotics based on a subjects age (Dominguez-Bello, Blaser, Ley, & Knight, 2011).

Similarly, prebiotics are an area of great interest owing to their associated benefits in health, including cancer inhibition, obesity and cholesterol reduction. Despite this, there appears to be little bioinformatic research solely dedicated to the study of prebiotics, with most studies focussing on pro and prebiotics collectively (see for example, Collison *et al.*, 2012). Nevertheless, a recent study using various bioinformatics techniques, such as phylogenetic microarrays, has been published where the gut microbiota was investigated after prebiotic administration in obese and diabetic mice (Everard *et al.*, 2011).

#### Food composition databases

Although not explored extensively by the bioinformatics community to date, it would be remiss of us not to discuss the various, notable, food composition database (FCDB) efforts that are ongoing globally. Such databases are important tools for nutritional assessment by health professionals (Pennington *et al.*, 2007) and are typically compiled on a national or regional basis (Scrimshaw, 1997). In the United States, the major food composition resource is the USDA National Nutrient Database for Standard Reference (NNDsr; <http://www.ars.usda.gov/ba/bhnrc/ndl>), which is a free, open access resource. The NNDsr is regularly updated and curated, and currently features data for more than 8000 foods making it one of the most comprehensive FCDBs. Accordingly, the NNDsr is utilised globally for nutritional assessment (Merchant and Dehghan, 2006). In Europe, EuroFIR (<http://www.eurofir.org/>) provides standardised FCDBs for food and nutritional scientists. Moreover, EuroFIR lists software developers among its target users, demonstrating the progression towards computational based research in this area. Although run by a non-profit organisation, this resource does require a membership subscription to gain access to data. Finally, the FAO/INFOODS Analytical Food Composition Database (<http://www.fao.org/infoods/infoods/tables-and-databases/>

faoinfoods-databases/en/) provides food composition data for foods that are commonly consumed globally.

While not specifically driven by bioinformaticians, FCDBs do subscribe to one of the central tenants of integrative bioinformatics in that there is a major concerted effort to standardise and assimilate data appropriately. Drivers of such initiatives include collaborative networks like INFOODS (Scrimshaw, 1997) and EuroFIR (http://www.eurofir.org/), while important historical initiatives include Eurofoods (Cheung, Wang, Ondetti, Sabo, & Cushman, 1980) and the European Food Consumption Survey Method (for more information see Vermeirssen, Van Der Bent, Van Camp, Van Amerongen, & Verstraete, 2004). The commendable work in this area serves as an important reference for food and nutritional sciences. Extension of these practices beyond food composition data could serve to greatly advance bioinformatic research in these domains.

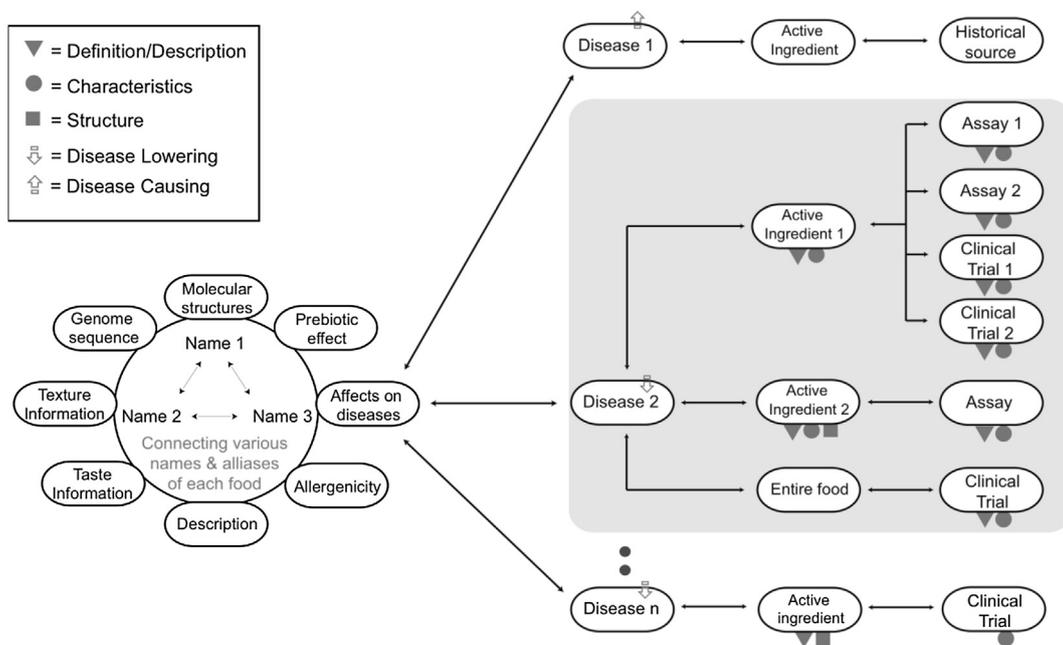
**FoodWikiDB: a consensus food information storage platform to facilitate bioinformatics**

Bioinformatics is a powerful means of quickly and accurately handling multiple data sources in high volume. However, a pre-requisite for the use of bioinformatics is the definition of data in a useable and uniform manner. Typically, data needs to be stored in a precise, fixed format or configuration in order for its expedient investigation by computational biology. Formatting and integrating food and nutritional data in such a manner allows for comparative analysis, which is essential for the discovery of new directions and opportunities.

The big picture: from a database structure to a user-friendly interface

One of the major problems preventing the use of bioinformatics in food and nutritional research is the absence of a consensus language or lexicon. For instance, the reporting of experimental data, detailing certain effects of a particular food, can be very disparate between different research laboratories, and can also vary greatly from food to food. Creating a uniform language is not an easy task but it is an achievable one, as we discuss below, and requires the initial step of arranging the data in an intuitive database-like representation. This will allow all types of food data to be deposited in a homogenous format, predefined in the database, facilitating the use of computational tools (see Fig. 2).

A brief examination of global monthly Google (conducted in January 2013) searches shows that searches for sentences containing the words food and cholesterol together amass to 5.1 million, while sentences containing the terms diabetes and food or blood sugar and food together accrue an incredible 12.5 million searches. Notably, these numbers are restricted to only English based searches. Accordingly, users are faced with the arduous task of filtering through thousands of websites to obtain accurate and reliable data. While a considerable amount of the information may be factual, there is currently no system in place to assess the credibility of the information retrieved. It is therefore necessary to define a common language and database structure that provides for user-friendly searches, while assuring the accuracy of the information retrieved.



**Fig. 2. Representation of the flow of information in the FoodWikiDB.**

The centre part of each piece of information in the database is represented by the name(s) of the food source(s) shown in a circle on the figure. The semi-transparent grey boxes contain examples of what can be stored for a given disease lowering effect of a given food source. The active ingredients can be known or not, and in some cases can be represented by an entire food.

Further to this, a wiki-like food database will provide a suitable platform for bioinformaticians to undertake comparative approaches between different data types and the prediction of novel experimental outcomes. For example, one can predict the result of an experiment where ingredient A is tested for activity against a given disease by simply deducing the results from another experiment where ingredient B, which is structurally homologous to A, was tested.

#### *What is a 'common language'?*

While there is pressure on the drug industry to adopt consensus methods and have global databases (such as PDB for 3-D structures), it is not the case for the food industry, where methods are free to vary from lab to lab. Consequently, it is difficult to arrive upon a generic consensus or database structure. This is demonstrated by the noticeable shortage of core databases for utilisation by the food industry and academics alike. To define a common language, a set of rules and attributes to classify an entity needs to be assembled. In our case, a food source is the entity, and the attributes that classify it could include, for example, structural and molecular information. Each defining element of the food should be stored separately, but all would be linked to the food source, as shown in Fig. 2.

While great strides in relation to a common language have already been made in the areas of food description and composition using lexicon systems such as LanguaL™ (Pennington & Butrum, 1991), AGROVOC ([www.fao.org/agrovoc/](http://www.fao.org/agrovoc/)) and INFOODS tagnames (Klensin, 1992), there is an apparent need for standardisation of these schemes (EFSA, 2011). Having a consensus language for all food and nutrition data will allow bioinformaticians to mine and compare information, to predict new leads, to detect inherent errors in data, and importantly, to guide experimental protocols required to validate concepts.

#### *Accurate scientific information presented in an accessible format*

One can imagine a FoodWikiDB resource, which will allow searches using a user-defined number of search terms of interest. The user will be provided with a webpage with the requested information linked to other pages containing a more detailed explanation of each point. For example, the webpage of a food source will contain information about this food, but will also link to all the ingredients (e.g., hydrolysates) that have been derived from this food, be they commercial or not. Each ingredient will in turn link to a separate page defining and describing its properties, which will additionally link out to other pages. The key feature of all this information is that the myriad of pages that a food source may link to is all nested under the same search term. These pages may also inter-link to the pages of other food sources if they share a common feature, for example, they were tested in the same assay.

The user can then choose to select only information supported by peer-reviewed sources. The user can refine or

broaden searches by entering exactly, or in more approximate terms, the information they require. Results could be presented in various drop box menus, where the user could choose all, or a more refined selection, depending on their needs.

#### *Logistics: structuring, populating, and querying the FoodWikiDB*

The wiki-like food database we are proposing here will address many of the above points. The FoodWikiDB should be accessible by all credible users (see 'Policing the content' section below), be it for adding information, or for querying its content. We propose the following structure as an initial framework for the FoodWikiDB:

#### *Structure outline for the data*

Fig. 2 shows an example of the FoodWikiDB layout for a given food source. Each food source (represented as a circle in Fig. 2) is identified by one or multiple names (i.e., scientific name or various common appellations) to allow for a broad user search scope, but these will be related to a single FoodWikiDB entry with a unique identifier. Each food source entry is linked to a number of elements. These include, for example, available data about the allergenicity of this food or its components, its genome sequence or that of its closest relatives, its taste and texture characteristics, its possible effects on probiotics, its potential prebiotic content, its enzyme content, and its implication in different diseases. Where possible each of these categories will be populated by links to established external websites and resources. This structure can be organised as displayed in Fig. 2. The food source will link out to a number of diseases that it has been shown to impact upon, with links provided to all pertinent experimental and clinical trial data (see Fig. 2). Here, there can be an option for the user to include widely used anecdotal remedies, should they wish, to supplement the scientifically proven data routinely returned in a search. In each case, known structures and properties of the ingredients or food can be retrieved (again where possible linking to established external resources). References to papers that discuss these results should also be provided. We believe that unpublished data can equally contain valuable information that may validate some experimental findings, therefore, this data should additionally be included under a non-published tag.

Very often, independent studies of the same ingredient or food show varying results, claiming a positive effect in one case, and conversely a negative in another. The FoodWikiDB would have the capacity to connect these studies together and highlight the major experimental differences between them (not shown in Fig. 2 for clarity reasons).

#### *Adding information to the database*

The FoodWikiDB, in line with the definition of a wiki, would be a dynamic database. Users around the world

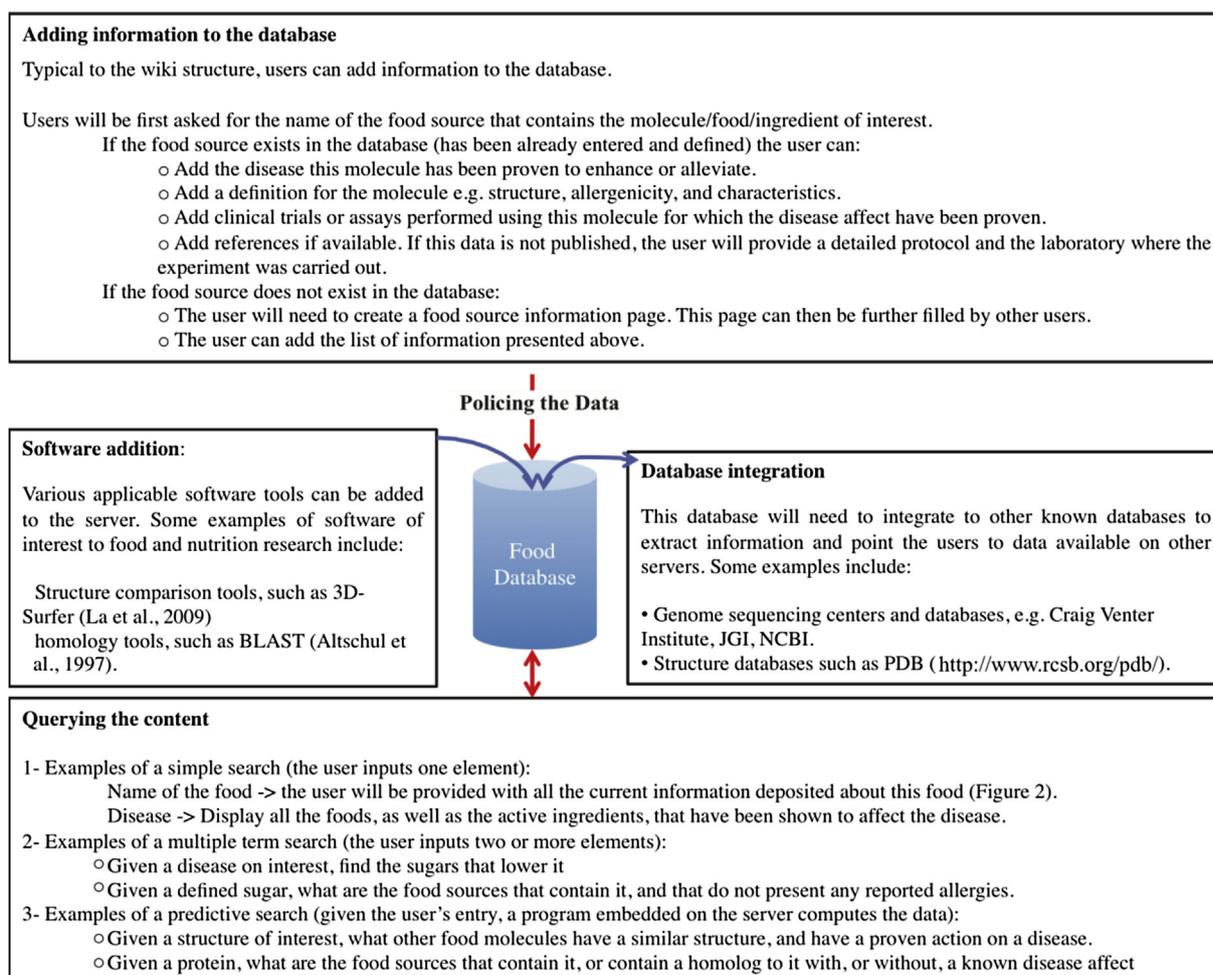
would be free to submit information about food and ingredients in the format specified above. New fields can be added to a food if they do not already exist or are not a pre-defined field. The user can fill out a form that will populate Fig. 2 or parts of it. Fig. 3 shows a schematic of how information can be added. When adding a particular ingredient, or an experimental assay, users will be required to relate these to a relevant pre-existing food source entry or entries. If an appropriate food source entry does not yet exist, the user will be prompted to create a new food source instance before adding any further information. This will ensure that new information will be linked to existing information in the database. We feel that the facility to add data, combined with the incorporation of links to information from external sources, will provide users with a truly comprehensive resource. Similar to integrative bioinformatic efforts for molecular data, novel data entered in the FoodWikiDB by a user would be subject to data integration protocols for

other relevant resources in order to uphold a culture of unified access.

### Extensions to the database

#### a. Data integration with external databases

There are many credible online resources that contain significant amounts of information about different food sources, as well as details about their genomes, proteins, sugars, structures, composition and other types of molecular data (Fig. 3). The FoodWikiDB would interact with these different resources to populate its content and would allow users to directly link to these external websites. Information about a given search item can be obtained from other databases using its associated attributes (e.g., accession number) as keys to perform indexing functionalities to retrieve the required information. Communication with



**Fig. 3. Representation of the different users and database administration interactions within the FoodWikiDB.**

Red arrows indicate a user's entry or query; blue arrows indicate additions made by the database administrators. Double-headed arrows indicate an exchange within the database. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article).

external databases can be facilitated through an Application Programming Interface (API).

#### b. Software additions

Enabling comparisons of the data stored in the FoodWikiDB and predicting new information would be of great interest to the community. To achieve this, various established software applications should be integrated into the database (Fig. 3). For example, a user may enter the sequence of a structure of interest and want to search the FoodWikiDB for similar structures associated with a particular health benefit. This would be possible with tools such as 3D-SURFER (La et al., 2009). Another use of the FoodWikiDB could be to search for peptides similar to one provided by the user, where homology searches, such as BLAST (Altschul et al., 1997), could be used to achieve this. Finally, a user could predict the enzymes used to create a particular hydrolysate of interest and further predict the different activities within this hydrolysate using tools such as EnzymePredictor (Vijayakumar et al., 2012). Many programs can be added to the FoodWikiDB to allow intelligent mining of the data. It is this feature that, we feel, will allow the use of bioinformatics to greatly advance food and nutritional research.

#### Policing the content

The data added to the FoodWikiDB needs to originate from a reliable source. It is thus important to police entries so that users can have confidence in the information they retrieve from the FoodWikiDB. One way to do this would be to require contributors to register a valid institutional email address as a means of authentication. Additionally, all users can be given the option to rate entries based on a similar rating scale to the Wikipedia article feedback tool.

#### Querying the content

To demonstrate some of the potential applications of the FoodWikiDB, we present the following select examples (represented in Fig. 3):

- A user can select a disease and query the database for associated relationships with food sources. The query could be tailored to limit results to only foods containing ingredients that have been clinically validated with published data. Further to this, the query can be tailored to select the subset of the returned food sources that have no related allergen effects, or the least amount of fat content.
- A user can choose to upload a structure and query the database for similar food ingredient structures with known, published, functions. They can also query the database to see if these similar structures have any information regarding their allergenicity.
- A user can search for foods with high oligosaccharide content.

- One can use known features of a given molecular function to predict novel potentially functional peptides. For example, the characteristics of angiotensin-converting enzyme (ACE) inhibitory peptides are known and well documented in the literature (Cheung et al., 1980; Vermeirssen et al., 2004). A user can choose to use this information to predict new bioactive peptides from a given food of interest found in the FoodWikiDB. By selecting a food and a prediction software implemented in the FoodWikiDB, the user can obtain a set of novel, predicted ACE-inhibitory peptides in this food.

The above examples of queries show how simple or complex the range of queries can be. This is provided for by the structure of the FoodWikiDB, which links the information together in an efficient and clear way.

#### Conclusion

Food and nutritional research, not unlike many other branches of life science, is fast becoming a very data rich discipline. As the move towards high throughput “omics” technologies perpetuates, the necessity of having appropriate means to curate and analyse such data volumes is paramount. The integration and implementation of bioinformatics is suitably poised to achieve this end. In this review, we have briefly outlined how the adoption of bioinformatics has impacted upon various field of research. By following this model, we feel that the capacities of food and nutritional research can be greatly enhanced. While undoubtedly, as detailed in this review, great strides have been made in computational based research in food and nutrition, we feel there is much scope for development in this area.

We feel that establishment of a centralised, comprehensive food and nutrition database is crucial to the progression of bioinformatics in this area. Here, we have specified the structural framework for this wiki-like food database, namely FoodWikiDB, as well as the intended content and functionality. Central to the formation of such a database will be the food and nutritional research community. This resource will depend heavily on the contribution of such researchers, and to ultimately succeed it will require their widespread adoption and utilisation. While this imposes a great challenge upon the members of these disciplines, we feel the outcome will not only enhance current research, but may also drive a paradigm shift in these sciences.

#### References

- Abee, T., Van Schaik, W., & Siezen, R. J. (2004). Impact of genomics on microbial food safety. *Trends in Biotechnology*, 22, 653–660.
- Ahmed, J., Preissner, S., Dunkel, M., Worth, C. L., Eckert, A., & Preissner, R. (2010). SuperSweet – a resource on natural and artificial sweetening agents. *Nucleic Acids Research*, 39, D377–D382.

- Altaier, E., Ramsay, S. L., Graber, A., Mewes, H.-W., Weinberger, K. M., & Suhre, K. (2008). Bioinformatics analysis of targeted metabolomics – uncovering old and new tales of diabetic mice under medication. *Endocrinology*, *149*, 3478–3489.
- Altschul, S. F., Madden, T. L., Schäffer, A. A., Zhang, J., Zhang, Z., Miller, W., et al. (1997). Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. *Nucleic Acids Research*, *25*, 3389–3402.
- Amiour, N., Merlino, M., Leroy, P., & Branlard, G. (2002). Proteomic analysis of amphiphilic proteins of hexaploid wheat kernels. *Proteomics*, *2*, 632–641.
- Baderschneider, B., Crevel, R. W. R., Earl, L. K., Lalljie, A., Sanders, D. J., & Sanders, I. J. (2002). Sequence analysis and resistance to pepsin hydrolysis as part of an assessment of the potential allergenicity of ice structuring protein type III HPLC 12. *Food and Chemical Toxicology*, *40*, 965–978.
- Behr, M. A., Wilson, M. A., Gill, W. P., Salamon, H., Schoolnik, G. K., Rane, S., et al. (1999). Comparative genomics of BCG vaccines by whole-genome DNA microarray. *Science*, *284*, 1520–1523.
- Bentley, S. D., & Parkhill, J. (2004). Comparative genomic structure of prokaryotes. *Annual Review of Genetics*, *38*, 771–792.
- Berman, H. M., Westbrook, J., Feng, Z., Gilliland, G., Bhat, T. N., Weissig, H., et al. (2000). The protein data bank. *Nucleic Acids Research*, *28*, 235–242.
- Biarnés, X., Marchiori, A., Giorgetti, A., Lanzara, C., Gasparini, P., Carloni, P., et al. (2010). Insights into the binding of phenyltiocarbamide (PTC) agonist to its target human TAS2R38 bitter receptor. *PLoS ONE*, *5*, e12394.
- Blundell, T. L., Sibanda, B. L., Montalvão, R. W., Brewerton, S., Chelliah, V., Worth, C. L., et al. (2006). Structural biology and bioinformatics in drug design: opportunities and challenges for target identification and lead discovery. *Philosophical Transactions of the Royal Society B: Biological Sciences*, *361*, 413–423.
- Brockhoff, A., Behrens, M., Niv, M. Y., & Meyerhof, W. (2010). Structural requirements of bitter taste receptor activation. *Proceedings of the National Academy of Sciences*, *107*, 11110–11115.
- Brul, S., Schuren, F., Montijn, R., Keijser, B. J. F., Van Der Spek, H., & Oomes, S. J. C. M. (2006). The impact of functional genomics on microbiological food quality and safety. *International Journal of Food Microbiology*, *112*, 195–199.
- Buchan, N. S., Rajpal, D. K., Webster, Y., Alatorre, C., Gudivada, R. C., Zheng, C., et al. (2011). The role of translational bioinformatics in drug discovery. *Drug Discovery Today*, *16*, 426–434.
- Bueschl, C., Kluger, B., Berthiller, F., Lirk, G., Winkler, S., Krska, R., et al. (2012). MetExtract: a new software tool for the automated comprehensive extraction of metabolite-derived LC/MS signals in metabolomics research. *Bioinformatics*, *28*, 736–738.
- Butte, A. J., & Ito, S. (2012). Translational bioinformatics: data-driven drug discovery and development. *Clinical Pharmacology & Therapeutics*, *91*, 949–952.
- Carlton, J. M., Adams, J. H., Silva, J. C., Bidwell, S. L., Lorenzi, H., Caler, E., et al. (2008). Comparative genomics of the neglected human malaria parasite *Plasmodium vivax*. *Nature*, *455*, 757–763.
- Cheung, H. S., Wang, F. L., Ondetti, M. A., Sabo, E. F., & Cushman, D. W. (1980). Binding of peptide substrates and inhibitors of angiotensin-converting enzyme. Importance of the COOH-terminal dipeptide sequence. *Journal of Biological Chemistry*, *255*, 401–407.
- Choi, H.-K., Mun, J.-H., Kim, D.-J., Zhu, H., Baek, J.-M., Mudge, J., et al. (2004). Estimating genome conservation between crop and model legume species. *Proceedings of the National Academy of Sciences of the United States of America*, *101*, 15289–15294.
- Chun, J., Grim, C. J., Hasan, N. A., Lee, J. H., Choi, S. Y., Haley, B. J., et al. (2009). Comparative genomics reveals mechanism for short-term and long-term clonal transitions in pandemic *Vibrio cholerae*. *Proceedings of the National Academy of Sciences*, *106*, 15442–15447.
- Cifuentes, A. (2009). Food analysis and foodomics. *Journal of Chromatography A*, *1216*, 7109.
- Colburn, W. A. (2003). Biomarkers in drug discovery and development: from target identification through drug marketing. *The Journal of Clinical Pharmacology*, *43*, 329–341.
- Collison, M., Hirt, R. P., Wipat, A., Nakjang, S., Sanseau, P., & Brown, J. R. (2012). Data mining the human gut microbiota for therapeutic targets. *Briefings in Bioinformatics*, *13*, 751–768.
- D'Alessandro, A., Zolla, L., & Scaloni, A. (2011). The bovine milk proteome: cherishing, nourishing and fostering molecular complexity. An interactomics and functional overview. *Molecular BioSystems*, *7*, 579–597.
- Demir, E., Babur, O., Dogrusoz, U., Gursoy, A., Nisanci, G., Cetin-Atalay, R., et al. (2002). Patika: an integrated visual environment for collaborative construction and analysis of cellular pathways. *Bioinformatics*, *18*, 996–1003.
- Desiere, F., German, B., Watzke, H., Pfeifer, A., & Saguy, S. (2001). Bioinformatics and data knowledge: the new frontiers for nutrition and foods. *Trends in Food Science & Technology*, *12*, 215–229.
- Dimitrov, I., Flower, D., & Doytchinova, I. (2013). AllerTOP – a server for in silico prediction of allergens. *BMC Bioinformatics*, *14*, S4.
- Dominguez-Bello, M. G., Blaser, M. J., Ley, R. E., & Knight, R. (2011). Development of the human gastrointestinal microbiota and insights from high-throughput sequencing. *Gastroenterology*, *140*, 1713–1719.
- Dong, D., Jones, G., & Zhang, S. (2009). Dynamic evolution of bitter taste receptor genes in vertebrates. *BMC Evolutionary Biology*, *9*, 12.
- Dotson, C. D., Zhang, L., Xu, H., Shin, Y.-K., Vignes, S., Ott, S. H., et al. (2008). Bitter taste receptors influence glucose homeostasis. *PLoS ONE*, *3*, e3974.
- Doumith, M., Cazalet, C., Simoes, N., Frangeul, L., Jacquet, C., Kunst, F., et al. (2004). New aspects regarding evolution and virulence of *Listeria monocytogenes* revealed by comparative genomics and DNA arrays. *Infection and Immunity*, *72*, 1072–1083.
- Dunkel, M., Schmidt, U., Struck, S., Berger, L., Gruening, B., Hossbach, J., et al. (2009). SuperScent – a database of flavors and scents. *Nucleic Acids Research*, *37*, D291–D294.
- EFSA. (2011). Report on the development of a Food Classification and Description System for exposure assessment and guidance on its implementation and use. *EFSA Journal*, *9*, 2489.
- Everard, A., Lazarevic, V., Derrien, M., Girard, M., Muccioli, G. G., Neyrinck, A. M., et al. (2011). Responses of gut microbiota and glucose and lipid metabolism to prebiotics in genetic obese and diet-induced leptin-resistant mice. *Diabetes*, *60*, 2775–2786.
- Fagerquist, C. K., Garbus, B. R., Miller, W. G., Williams, K. E., Yee, E., Bates, A. H., et al. (2010). Rapid identification of protein biomarkers of *Escherichia coli* O157:H7 by matrix-assisted laser desorption ionization-time-of-flight–time-of-flight mass spectrometry and top-down proteomics. *Analytical Chemistry*, *82*, 2717–2725.
- Fagerquist, C. K., Garbus, B. R., Williams, K. E., Bates, A. H., Boyle, S., & Harden, L. A. (2009). Web-based software for rapid top-down proteomic identification of protein biomarkers, with implications for bacterial identification. *Applied and Environmental Microbiology*, *75*, 4341–4353.
- Fang, H., Xu, J., Ding, D., Jackson, S., Patel, I., Frye, J., et al. (2011). An FDA bioinformatics tool for microbial genomics research on molecular characterization of bacterial foodborne pathogens using microarrays. *BMC Bioinformatics*, *11*, S4.
- FAO/WHO. (2001). *Evaluation of allergenicity of genetically modified foods. Report of a joint FAO/WHO expert consultation on*

- allergenicity of foods derived from biotechnology*. Rome: Food and Agriculture Organization of the United Nations (FAO).
- Fedorova, N. D., Khaldi, N., Joardar, V. S., Maiti, R., Amedeo, P., Anderson, M. J., et al. (2008). Genomic islands in the pathogenic filamentous fungus *Aspergillus fumigatus*. *PLoS Genetics*, *4*, e1000046.
- Fei, Z., Tang, X., Alba, R. M., White, J. A., Ronning, C. M., Martin, G. B., et al. (2004). Comprehensive EST analysis of tomato and comparative genomics of fruit ripening. *The Plant Journal*, *40*, 47–59.
- Fiers, M., Kleter, G., Nijland, H., Peijnenburg, A., Nap, J., & Van Ham, R. (2004). Allermatch<sup>TM</sup>, a webtool for the prediction of potential allergenicity according to current FAO/WHO Codex alimentarius guidelines. *BMC Bioinformatics*, *5*, 133.
- Fuller, R. (1989). Probiotics in man and animals. *The Journal of Applied Bacteriology*, *66*, 365–378.
- García-Gimeno, R. M. A., Hervás-Martínez, C., & De Silónizc, M. I. (2002). Improving artificial neural networks with a pruning methodology and genetic algorithms for their application in microbial growth prediction in food. *International Journal of Food Microbiology*, *72*, 19–30.
- Gaulton, A., Bellis, L. J., Bento, A. P., Chambers, J., Davies, M., Hersey, A., et al. (2012). ChEMBL: a large-scale bioactivity database for drug discovery. *Nucleic Acids Research*, *40*, D1100–D1107.
- Gibbs, B. F., Zougman, A., Masse, R., & Mulligan, C. (2004). Production and characterization of bioactive peptides from soy hydrolysate and soy-fermented food. *Food Research International*, *37*, 123–131.
- Gieger, C., Geistlinger, L., Altmaier, E., Hrabé De Angelis, M., Kronenberg, F., Meitinger, T., et al. (2008). Genetics meets metabolomics: a genome-wide association study of metabolite profiles in human serum. *PLoS Genetics*, *4*, e1000282.
- Hajduk, P. J., Huth, J. R., & Tse, C. (2005). Predicting protein druggability. *Drug Discovery Today*, *10*, 1675–1682.
- Hartmann, R., & Meisel, H. (2007). Food-derived peptides with biological activity: from research to food applications. *Current Opinion in Biotechnology*, *18*, 163–169.
- Herrero, M., García-Cañas, V., Simo, C., & Cifuentes, A. (2010). Recent advances in the application of capillary electromigration methods for food analysis and foodomics. *Electrophoresis*, *31*, 205–228.
- Hileman, R. E., Silvanovich, A., Goodman, R. E., Rice, E. A., Holleschak, G., Astwood, J. D., et al. (2002). Bioinformatic methods for allergenicity assessment using a comprehensive allergen database. *International Archives of Allergy and Immunology*, *128*, 280–291.
- Hopkins, A. L., & Groom, C. R. (2002). The druggable genome. *Nature Reviews Drug Discovery*, *1*, 727–730.
- Huang, A. L., Chen, X., Hoon, M. A., Chandrashekar, J., Guo, W., Trankner, D., et al. (2006). The cells and logic for mammalian sour taste detection. *Nature*, *442*, 934–938.
- Illig, T., Gieger, C., Zhai, G., Romisch-Margl, W., Wang-Sattler, R., Prehn, C., et al. (2010). A genome-wide perspective of genetic variation in human metabolism. *Nature Genetics*, *42*, 137–141.
- Ivanciuc, O., Garcia, T., Torres, M., Schein, C. H., & Braun, W. (2009). Characteristic motifs for families of allergenic proteins. *Molecular Immunology*, *46*, 559–568.
- Ivanciuc, O., Schein, C. H., & Braun, W. (2003). SDAP: database and computational tools for allergenic proteins. *Nucleic Acids Research*, *31*, 359–362.
- Jenkins, J. A., Griffiths-Jones, S., Shewry, P. R., Breiteneder, H., & Mills, E. N. C. (2005). Structural relatedness of plant food allergens with specific reference to cross-reactive allergens: an in silico analysis. *Journal of Allergy and Clinical Immunology*, *115*, 163–170.
- Katajamaa, M., Miettinen, J., & Orešič, M. (2006). MZmine: toolbox for processing and visualization of mass spectrometry based molecular profile data. *Bioinformatics*, *22*, 634–636.
- Khaldi, N. (2012). Bioinformatics approaches for identifying new therapeutic bioactive peptides in food. *Functional Foods in Health and Disease*, *2*, 325–338.
- Khaldi, N., Collemare, J., Lebrun, M. H., & Wolfe, K. H. (2008). Evidence for horizontal transfer of a secondary metabolite gene cluster between fungi. *Genome Biology*, *9*, R18.
- Khaldi, N., Seifuddin, F. T., Turner, G., Haft, D., Nierman, W. C., Wolfe, K. H., et al. (2010). SMURF: genomic mapping of fungal secondary metabolite clusters. *Fungal Genetics and Biology*, *47*, 736–741.
- Khaldi, N., & Wolfe, K. H. (2008). Elusive origins of the extra genes in *Aspergillus oryzae*. *PLoS ONE*, *3*, e3036.
- Khaldi, N., & Wolfe, K. H. (2011). Evolutionary origins of the fumonisin secondary metabolite gene cluster in *Fusarium verticillioides* and *Aspergillus niger*. *International Journal of Evolutionary Biology*, *2011*, 423821.
- Klensin, J. C. (1992). *INFOODS food composition data interchange handbook*. United Nations University Press Tokyo.
- Korhonen, H., & Pihlanto, A. (2006). Bioactive peptides: production and functionality. *International Dairy Journal*, *16*, 945–960.
- La, D., Esquivel-Rodríguez, J., Venkatraman, V., Li, B., Sael, L., Ueng, S., et al. (2009). 3D-SURFER: software for high-throughput protein surface comparison and analysis. *Bioinformatics*, *25*, 2843–2844.
- Lemay, D. G., Zivkovic, A. M., & German, J. B. (2007). Building the bridges to bioinformatics in nutrition research. *The American Journal of Clinical Nutrition*, *86*, 1261–1269.
- Levit, A., Barak, D., Behrens, M., Meyerhof, W., Niv, M. Y., Vaidehi, N., et al. (2012). Homology model-assisted elucidation of binding sites in GPCRs. In N. Vaidehi, & J. Klein-Seetharaman (Eds.), *Membrane protein structure and dynamics: Methods and protocols*. Humana Press.
- Lienau, E. K., Strain, E., Wang, C., Zheng, J., Ottesen, A. R., Keys, C. E., et al. (2011). Identification of a salmonellosis outbreak by means of molecular sequencing. *New England Journal of Medicine*, *364*, 981–982.
- Lin, J., Bruni, F. M., Fu, Z., Maloney, J., Bardina, L., Boner, A. L., et al. (2012). A bioinformatics approach to identify patients with symptomatic peanut allergy using peptide microarray immunoassay. *Journal of Allergy and Clinical Immunology*, *129*, 1321–1328.e5.
- Liu, M., Nauta, A., Francke, C., & Siezen, R. J. (2008). Comparative genomics of enzymes in flavor-forming pathways from amino acids in lactic acid bacteria. *Applied and Environmental Microbiology*, *74*, 4590–4600.
- Lommen, A. (2009). MetAlign: interface-driven, versatile metabolomics tool for hyphenated full-scan mass spectrometry data preprocessing. *Analytical Chemistry*, *81*, 3079–3086.
- Lynn, D. J., Higgs, R., Gaines, S., Tierney, J., James, T., Lloyd, A. T., et al. (2004). Bioinformatic discovery and initial characterisation of nine novel antimicrobial peptide genes in the chicken. *Immunogenetics*, *56*, 170–177.
- Lynn, D. J., Lloyd, A. T., & O'Farrelly, C. (2003). In silico identification of components of the Toll-like receptor (TLR) signaling pathway in clustered chicken expressed sequence tags (ESTs). *Veterinary Immunology and Immunopathology*, *93*, 177–184.
- Ma, L.-J., Van Der Does, H. C., Borkovich, K. A., Coleman, J. J., Daboussi, M.-J., Di Pietro, A., et al. (2010). Comparative genomics reveals mobile pathogenicity chromosomes in *Fusarium*. *Nature*, *464*, 367–373.
- Magrane, M., & Consortium, U. (2011). UniProt Knowledgebase: a hub of integrated protein data. *Database*, *2011*, bar009.
- Majumder, K., & Wu, J. (2010). A new approach for identification of novel antihypertensive peptides from egg proteins by QSAR and bioinformatics. *Food Research International*, *43*, 1371–1378.

- Makarenkov, V., Zentilli, P., Kevorkov, D., Gagarin, A., Malo, N., & Nadon, R. (2007). An efficient method for the detection and elimination of systematic error in high-throughput screening. *Bioinformatics*, *23*, 1648–1657.
- Makarova, K., Slesarev, A., Wolf, Y., Sorokin, A., Mirkin, B., Koonin, E., et al. (2006). Comparative genomics of the lactic acid bacteria. *Proceedings of the National Academy of Sciences*, *103*, 15611–15616.
- Mari, A., Scala, E., Palazzo, P., Ridolfi, S., Zennaro, D., & Carabella, G. (2006). Bioinformatics applied to allergy: allergen databases, from collecting sequence information to data integration. The allergome platform as a model. *Cellular Immunology*, *244*, 97–100.
- Matsui, T., Li, C.-H., & Osajima, Y. (1999). Preparation and characterization of novel bioactive peptides responsible for angiotensin I-converting enzyme inhibition from wheat germ. *Journal of Peptide Science*, *5*, 289–297.
- Mayer, K. F. X., Martis, M., Hedley, P. E., Šimková, H., Liu, H., Morris, J. A., et al. (2011). Unlocking the barley genome by chromosomal and comparative genomics. *The Plant Cell Online*, *23*, 1249–1263.
- Mazière, P., & Enright, A. J. (2007). Prediction of microRNA targets. *Drug Discovery Today*, *12*, 452–458.
- McInnes, C. (2007). Virtual screening strategies in drug discovery. *Current Opinion in Chemical Biology*, *11*, 494–502.
- Merchant, A., & Dehghan, M. (2006). Food composition database development for between country comparisons. *Nutrition Journal*, *5*, 2.
- Mery, D., Pedreschi, F., & Soto, A. (2012). Automated design of a computer vision system for visual food quality evaluation. *Food and Bioprocess Technology*, 1–16.
- Mochida, K., & Shinozaki, K. (2010). Genomics and bioinformatics resources for crop improvement. *Plant and Cell Physiology*, *51*, 497–523.
- Mooney, C., Haslam, N. J., Holton, T. A., Pollastri, G., & Shields, D. C. (2013). PeptideLocator: prediction of bioactive peptides in protein sequences. *Bioinformatics*, *29*, 1120–1126.
- Mooney, C., Haslam, N. J., Pollastri, G., & Shields, D. C. (2012). Towards the improved discovery and design of functional peptides: common features of diverse classes permit generalized prediction of bioactivity. *PLoS ONE*, *7*, e45012.
- Mullally, M. M., Meisel, H., & Fitzgerald, R. J. (1997). Identification of a novel angiotensin-I-converting enzyme inhibitory peptide corresponding to a tryptic fragment of bovine beta-lactoglobulin. *FEBS Letters*, *402*, 99–101.
- Ogura, Y., Ooka, T., Iguchi, A., Toh, H., Asadulghani, M., Oshima, K., et al. (2009). Comparative genomics reveal the mechanism of the parallel evolution of O157 and non-O157 enterohemorrhagic *Escherichia coli*. *Proceedings of the National Academy of Sciences*, *106*, 17939–17944.
- Omelchenko, M., Wolf, Y., Gaidamakova, E., Matrosova, V., Vasilenko, A., Zhai, M., et al. (2005). Comparative genomics of *Thermus thermophilus* and *Deinococcus radiodurans*: divergent routes of adaptation to thermophily and radiation resistance. *BMC Evolutionary Biology*, *5*, 57.
- Orešič, M. (2009). Metabolomics, a novel tool for studies of nutrition, metabolism and lipid dysfunction. *Nutrition, Metabolism and Cardiovascular Diseases*, *19*, 816–824.
- Pampanin, D. M., Larssen, E., Provan, F., Sivertsvik, M., Ruoff, P., & Sydnes, M. O. (2012). Detection of small bioactive peptides from Atlantic herring (*Clupea harengus* L.). *Peptides*, *34*, 423–426.
- Paterson, A. H., Bowers, J. E., & Chapman, B. A. (2004). Ancient polyploidization predating divergence of the cereals, and its consequences for comparative genomics. *Proceedings of the National Academy of Sciences of the United States of America*, *101*, 9903–9908.
- Paterson, A. H., Bowers, J. E., Feltus, F. A., Tang, H., Lin, L., & Wang, X. (2009). Comparative genomics of grasses promises a bountiful harvest. *Plant Physiology*, *149*, 125–131.
- Pennington, J. A. T., & Butrum, R. R. (1991). Food descriptions using taxonomy and the 'Languag' system. *Trends in Food Science & Technology*, *2*, 285–288.
- Pennington, J. A. T., Stumbo, P. J., Murphy, S. P., Mcnutt, S. W., Eldridge, A. L., McCabe-Sellers, B. J., et al. (2007). Food composition data: the foundation of dietetic practice and research. *Journal of the American Dietetic Association*, *107*, 2105–2113.
- Pridmore, R. D., Berger, B., Desiere, F., Vilanova, D., Barretto, C., Pittet, A.-C., et al. (2004). The genome sequence of the probiotic intestinal bacterium *Lactobacillus johnsonii* NCC 533. *Proceedings of the National Academy of Sciences of the United States of America*, *101*, 2512–2517.
- Qian, W., Jia, Y., Ren, S.-X., He, Y.-Q., Feng, J.-X., Lu, L.-F., et al. (2005). Comparative and functional genomic analyses of the pathogenicity of phytopathogen *Xanthomonas campestris* pv. *campestris*. *Genome Research*, *15*, 757–767.
- Rath, C., Alexandrov, T., Higginbottom, S., Song, J., Milla, M., Fischbach, M., et al. (2012). Molecular analysis of model gut microbiotas by imaging mass spectrometry and nano-desorption electrospray ionization reveals dietary metabolite transformations. *Analytical Chemistry*, *84*, 9259–9267.
- Rubin, E. M. (2008). Genomics of cellulosic biofuels. *Nature*, *454*, 841–845.
- Rubin, G. M., Yandell, M. D., Wortman, J. R., Gabor Miklos, G. L., Nelson, C. R., Hariharan, I. K., et al. (2000). Comparative genomics of the eukaryotes. *Science*, *287*, 2204–2215.
- Saha, S., & Raghava, G. P. S. (2006). AlgPred: prediction of allergenic proteins and mapping of IgE epitopes. *Nucleic Acids Research*, *34*, W202–W209.
- Scrimshaw, N. S. (1997). INFOODS: the international network of food data systems. *The American Journal of Clinical Nutrition*, *65*, 1190S–1193S.
- Shi, P., & Zhang, J. (2006). Contrasting modes of evolution between vertebrate sweet/umami receptor genes and bitter receptor genes. *Molecular Biology and Evolution*, *23*, 292–300.
- Taboureau, O., Baell, J. B., Fernández-Recio, J., & Villoutreix, B. O. (2012). Established and emerging trends in computational drug discovery in the structural genomics era. *Chemistry & Biology*, *19*, 29–41.
- Talevi, A., Enrique, A. V., & Bruno-Blanch, L. E. (2012). Anticonvulsant activity of artificial sweeteners: a structural link between sweet-taste receptor T1R3 and brain glutamate receptors. *Bioorganic & Medicinal Chemistry Letters*, *22*, 4072–4074.
- Taylor, R. D., Jewsbury, P. J., & Essex, J. W. (2002). A review of protein-small molecule docking methods. *Journal of Computer-Aided Molecular Design*, *16*, 151–166.
- Torrent, M., Andreu, D., Nogués, V. M., & Boix, E. (2011). Connecting peptide physicochemical and antimicrobial properties by a rational prediction model. *PLoS ONE*, *6*, e16968.
- Torrent, M., Nogués, V., & Boix, E. (2009). A theoretical approach to spot active regions in antimicrobial proteins. *BMC Bioinformatics*, *10*, 373.
- Turroni, F., Marchesi, J. R., Foroni, E., Gueimonde, M., Shanahan, F., Margolles, A., et al. (2009). Microbiomic analysis of the bifidobacterial population in the human distal gut. *ISME Journal*, *3*, 745–751.
- Van Kranenburg, R., Kleerebezem, M., Van Hylckama Vlieg, J., Ursing, B. R. M., Boekhorst, J., Smit, B. A., et al. (2002). Flavour formation from amino acids by lactic acid bacteria: predictions from genome sequence analysis. *International Dairy Journal*, *12*, 111–121.
- Varshney, R. K., Graner, A., & Sorrells, M. E. (2005). Genomics-assisted breeding for crop improvement. *Trends in Plant Science*, *10*, 621–630.

- Ventura, M., Turrioni, F., Zomer, A., Foroni, E., Giubellini, V., Bottacini, F., et al. (2009). The *Bifidobacterium dentium* Bd1 genome sequence reflects its genetic adaptation to the human oral cavity. *PLoS Genetics*, *5*, e1000785.
- Vercruyse, L., Smagghe, G., Van Der Bent, A., Van Amerongen, A., Ongenaert, M., & Van Camp, J. (2009). Critical evaluation of the use of bioinformatics as a theoretical tool to find high-potential sources of ACE inhibitory peptides. *Peptides*, *30*, 575–582.
- Vermeirssen, V., Van Camp, J., & Verstraete, W. (2002). Optimisation and validation of an angiotensin-converting enzyme inhibition assay for the screening of bioactive peptides. *Journal of Biochemical and Biophysical Methods*, *51*, 75–87.
- Vermeirssen, V., Van Der Bent, A., Van Camp, J., Van Amerongen, A., & Verstraete, W. (2004). A quantitative in silico analysis calculates the angiotensin I converting enzyme (ACE) inhibitory activity in pea and whey protein digests. *Biochimie*, *86*, 231–239.
- Vijayakumar, V., Guerrero, A. N., Davey, N., Lebrilla, C. B., Shields, D. C., & Khaldi, N. (2012). EnzymePredictor: a tool for predicting and visualizing enzymatic cleavages of digested proteins. *Journal of Proteome Research*, *11*, 6056–6065.
- Wassarman, K. M., Repoila, F., Rosenow, C., Storz, G., & Gottesman, S. (2001). Identification of novel small RNAs using comparative genomics and microarrays. *Genes & Development*, *15*, 1637–1651.
- Wiener, A., Shudler, M., Levit, A., & Niv, M. Y. (2012). BitterDB: a database of bitter compounds. *Nucleic Acids Research*, *40*, D413–D419.
- Wishart, D. S., Knox, C., Guo, A. C., Shrivastava, S., Hassanali, M., Stothard, P., et al. (2006). DrugBank: a comprehensive resource for in silico drug discovery and exploration. *Nucleic Acids Research*, *34*, D668–D672.
- Wohlbach, D. J., Kuo, A., Sato, T. K., Potts, K. M., Salamov, A. A., Labutti, K. M., et al. (2011). Comparative genomics of xylose-fermenting fungi for enhanced biofuel production. *Proceedings of the National Academy of Sciences*, *108*, 13212–13217.
- Wooding, S., Bufe, B., Grassi, C., Howard, M. T., Stone, A. C., Vazquez, M., et al. (2006). Independent evolution of bitter-taste sensitivity in humans and chimpanzees. *Nature*, *440*, 930–934.

# ScienceDirect

## Essential to Research Success

Discover ScienceDirect today and access scientific, technical and medical content from:

- More than **12 million articles** and book chapters
- Close to **3,000 journal titles** and **20,000 eBooks**
- A growing number of freely available **open access** journals
- **Digital archives** that reach as far back as 1823



Search with confidence across:

- Physical Sciences & Engineering
- Life Sciences
- Health Sciences
- Social Sciences & Humanities



[www.info.sciencedirect.com](http://www.info.sciencedirect.com)

Register now to take advantage of customized features and services.