

# Tools for analyzing and mapping scholarly publications not indexed by the Web of Science: the case of fruit and vegetable publications by the French National Institute for Agricultural Research (INRA)

Marie Violaine TATRY<sup>1</sup>, Dominique FOURNIER<sup>1</sup>, Benoît JEANNEQUIN<sup>2</sup>, Françoise DOSBA<sup>3\*</sup>

<sup>1</sup> INRA, Unité doc. ERIST,  
F-34060 Montpellier, France

<sup>2</sup> INRA, Domaine Exp. INRA  
SAD, Le Mas Blanc,  
F-66200 Alénia, France

<sup>3</sup> Montpellier SupAgro, INRA,  
UMR Agap – Amélioration  
génétique et adaptation  
de la plante,  
F-34060 Montpellier, France  
dosbaf@supagro.inra.fr

## Tools for analyzing and mapping scholarly publications not indexed by the Web of Science: the case of fruit and vegetable publications by the French National Institute for Agricultural Research (INRA).

**Abstract — Introduction.** The fruit and vegetable sector group of the French National Institute for Agricultural Research (INRA) carried out a bibliometric analysis of INRA publications on fruits and vegetables produced between 2002 and 2006. While a previous study focused on scientific publications indexed by the Web of Science, this study took into account the remaining INRA publications, *i.e.*, gray literature, technical papers and scientific papers not indexed by the Web of Science. **Materials and methods.** Different steps were taken to constitute and process a corpus of INRA researchers' publication data. The citation corpus gathering bibliographic data was mainly constituted using the INRA institutional repository (ProdINRA). To ensure exhaustivity, a survey was carried out to collect INRA researchers' publications. Due to the different types of publications involved, we had to develop an original method to constitute a complete database which could allow bibliometric analysis. The Sphinx Plus<sup>2</sup> Survey program was used to analyze publications and Pajek software was used to visualize networks. **Results.** A total of 2,224 publications were authored by INRA between 2002 and 2006. We analyzed the typology of fruits and vegetables studied at INRA, the INRA research centers and research divisions involved, the main research topics, and the national and international partners of INRA authors. Graphic visualizations allowed an attractive presentation of relationships between authors within and outside INRA. **Conclusion.** This study led to the mapping of INRA fruit and vegetable research publications. It confirmed that INRA plays an active role in the field of fruit and vegetable research. French research is carried out in the main INRA research centers located in horticultural and/or fruit regions, particularly in South-Eastern France. Despite the fact that the nature of these publications provides less opportunity for co-publication, this study highlights the relationships between INRA and French partners or foreign countries.

France / fruits / vegetables / research institutions / scientific production / data collection / information science / information processing / bibliometrics / partnerships

## Des outils pour analyser et cartographier des publications scientifiques non indexées dans le Web of Science : exemple des publications de l'Institut national de la recherche agronomique (INRA) sur les fruits et légumes.

**Résumé — Introduction.** Le groupe filière fruits et légumes de l'Institut national de la recherche agronomique (INRA) a réalisé une étude bibliométrique des publications de l'INRA sur les fruits et légumes au cours de la période 2002–2006. Une précédente étude avait été consacrée aux publications indexées dans le Web of Science, l'étude présentée ici prend en compte les publications INRA restantes, c'est-à-dire la littérature grise, les articles dans des revues techniques et les articles scientifiques non indexés dans le Web of Science. **Matériel et méthodes.** Différentes étapes ont permis de constituer et traiter un corpus des publications des chercheurs de l'INRA. Le corpus a été constitué en interrogeant l'archive institutionnelle de l'INRA (ProdINRA). Afin d'assurer l'exhaustivité des données, une enquête a été réalisée pour collecter les publications des chercheurs INRA. Du fait des différents types de publications étudiés, une méthode originale a dû être développée pour constituer une base de données permettant une analyse bibliométrique. Le logiciel Sphinx Plus<sup>2</sup> a été utilisé pour l'analyse des publications et le logiciel Pajek a été utilisé pour visualiser les réseaux. **Résultats.** Nous avons identifié 2 224 publications INRA sur la période 2002–2006. Ces publications ont été analysées afin de repérer les espèces de fruits et de légumes étudiés à l'INRA, les centres et les départements de recherche impliqués, les thématiques de recherche ainsi que les partenaires français et internationaux de l'INRA. L'utilisation de représentation de réseaux a permis une visualisation des relations entre acteurs (INRA et hors INRA). **Conclusion.** Cette étude a permis de cartographier les publications de l'INRA sur les fruits et les légumes. Le rôle actif de l'INRA dans la recherche sur les fruits et légumes a été confirmé. En France, cette recherche est menée dans les centres INRA au cœur des principales régions horticolaes et/ou fruitières, notamment dans le sud-est de la France. Même si les publications étudiées donnent assez peu lieu à des co-publications, cette étude montre les relations entretenues par l'INRA avec les institutions françaises, les partenaires de la filière et les pays étrangers.

France / fruits / légumes / institutions de recherche / publication scientifique / collecte de données / science de l'information / traitement de l'information / bibliométrie / partenariat

\* Correspondence and reprints

Fruits, 2011, vol. 66, p. 131–140  
© 2011 Cirad/EDP Sciences  
All rights reserved  
DOI: 10.1051/fruits/2011022  
www.fruits-journal.org

RESUMEN ESPAÑOL, p. 140

## 1. Introduction

The fruit and vegetable sector group of the French National Institute for Agricultural Research (INRA) undertook a bibliometric analysis of INRA's scientific output on fruits and vegetables.

A first bibliometric study was carried out on Web of Science (WoS) publications, covering the 2000 to 2006 period. The results made it possible to identify the main fruit and vegetable species focused on by INRA research, and the INRA research centers and INRA research divisions involved in fruit and vegetable research. INRA's collaboration network was also taken into account. This first study led to a report, an executive summary [1] and an article [2].

As INRA is a mission-oriented research institute, its researchers produce and disseminate knowledge not only through articles indexed by the WoS, but also through conference proceedings, technical papers, books and book chapters. They furthermore supervise PhD theses, which are considered important scientific outputs. It consequently appeared important to the INRA fruit and vegetable sector group to carry out a complementary study on all of INRA's scholarly publications which are not indexed by the WoS. In this study, we shall refer to these references as non-WoS publications.

Our paper describes the main tools used for this second part of the bibliometric analysis and presents the main results: (1) the different document types in the corpus, (2) the fruit and vegetable species studied, (3) the main subjects studied, (4) the INRA research centers and divisions involved, and

(5) INRA's domestic and international collaborations.

## 2. Methodology

Our analysis was performed on non-WoS publications disseminated from 2002 to 2006.

The data processing chain was composed of five steps (*figure 1*):

– Two different methods were used to retrieve INRA fruit and vegetable publications: (1) a search was carried out in the institutional repository ProdINRA, and (2) INRA research units (identified in the previous WoS bibliometric analysis) were contacted for their publication lists.

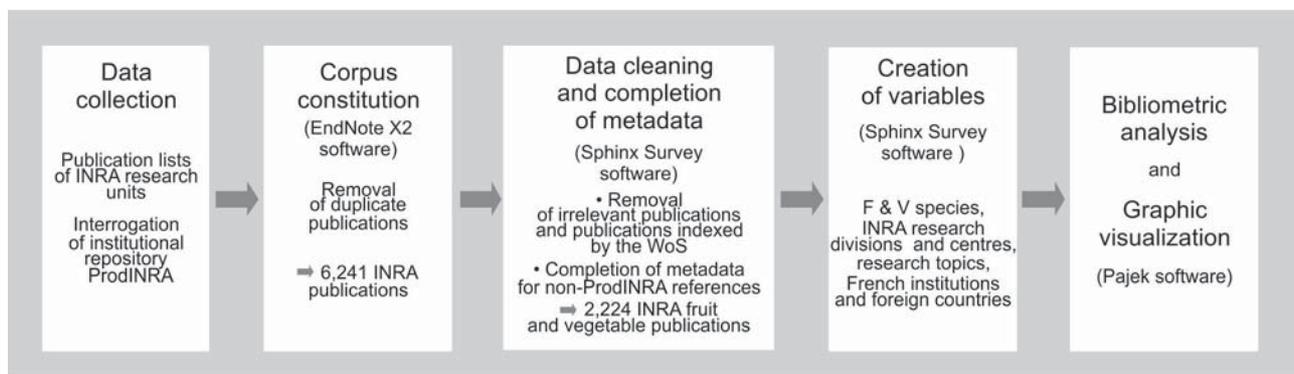
– Publications were managed with the reference management software Endnote X2, and duplicate references were removed.

– Data were cleaned, *i.e.*, irrelevant publications or publications indexed by the WoS were removed and metadata were completed, in particular authors' addresses. Sphinx Survey software was used for this step.

– New variables were created from bibliographic reference fields: the 'Title', 'Keywords' and 'Abstract' fields were used to code the species and subjects studied. An 'Address' field was used to code INRA research divisions and centers, other French institutions, and foreign countries. Sphinx Survey software, Lexica option, was used for this step.

– Data analysis was performed using Sphinx Survey software. Networks were drawn

**Figure 1.** Fruit and vegetable data processing chain used for studying the non-WoS publications of INRA's authors (France).



using Pajek software, which allows analysis and graphic visualization [3].

As we wished to measure participation in an article, we used full integer counts [4]. For example, an article co-signed by Spanish and French authors was credited with a unitary participation for Spain and for France.

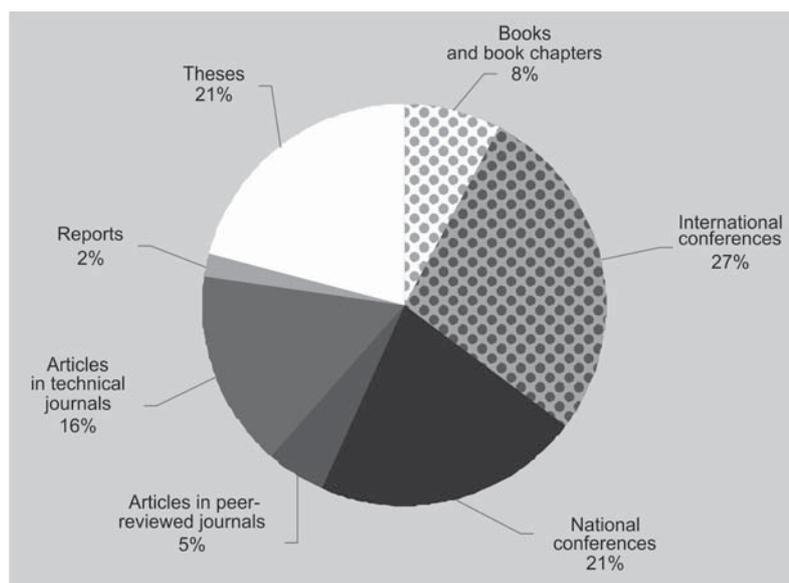
### 3. Main findings

#### 3.1. Description of the corpus

During the 2002–2006 period, INRA authors published 3,306 publications on fruits and vegetables (F&V). Of these 3,306 publications, 1,082 were indexed by the WoS and 2,224 publications were not. These non-WoS publications represented 67% of INRA publications. This confirms that non-WoS publications are important outputs for a mission-oriented research institute such as INRA. We will refer henceforth to the 2,224 non-WoS publications as the ‘F&V corpus’.

The different document types of the non-WoS publications (figure 2) were:

- international conferences (27% of the F&V corpus): the main conference organizers were the ISHS (International Society for Horticultural Science) and FESPb (Federation of European Societies of Plant Biology),
- national conferences (21% of the F&V corpus),
- peer-reviewed journals not indexed by the WoS (5% of the F&V corpus): most of the articles were published in *Fruits* (not indexed by the WoS during the period studied), *Cahiers Agricultures* (not indexed by the WoS during the period studied), *EPPO Bulletin*, and *Economies et sociétés*,
- technical journals (16% of the F&V corpus): the two main technical journals are *Réussir Fruits et Légumes* and *Phytoma la Défense des Végétaux*,
- books (3% of the F&V corpus) and book chapters (5% of the F&V corpus); according to the publishers, the books are aimed at the scientific community (Kluwer Academic Press, CRC Press, CABI International,



**Figure 2.**

Distribution of the document types of the non-WoS publications of INRA's authors (France), regarding the fruit and vegetable corpus (2,224 publications) over the 2002–2006 period.

Lavoisier) and economic actors of the fruit and vegetable chain (Ctifl),

- theses (21% of the F&V corpus): PhD theses and master degree theses, respectively, accounted for 3% and 9% of the F&V corpus; various other theses accounted for the remaining 9%,
- reports (2% of the F&V corpus).

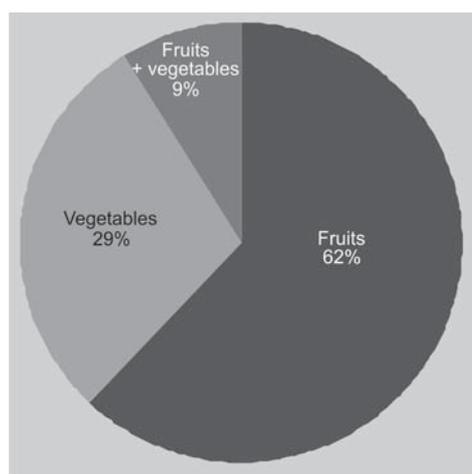
As a mission-oriented research institute, INRA disseminates knowledge to F&V stakeholders through national conferences, technical journals and technical books published by economic actors (e.g., Ctifl). International conferences, peer-reviewed journals, scientific books published by scientific editors (Kluwer Academic Press, CRC Press, CABI International, Lavoisier) and theses are aimed at researchers.

#### 3.2. The typology of fruits and vegetables studied at INRA

To identify the species studied at INRA, a text analysis was applied to the ‘Title’, ‘Abstract’, and ‘Keywords’ fields using the list of F&V species used in the first bibliometric analysis [5].

INRA publications covered 24 different fruit species and 27 different vegetable species.

**Figure 3.** Distribution of the non-WoS publications of INRA's authors (France) by fruit (F) and/or vegetable (V) species (F&V corpus: 2,224 publications, 2002–2006).



In the 'F&V corpus', fruit species only were the subject of 62% of the publications, vegetable species only of 29%, and both fruit and vegetable species of 9% (*figure 3*). We shall refer to publications related to fruit species as the 'F corpus' and to vegetable species as the 'V corpus'.

The twelve main fruit species or groups of species studied by INRA represent 81% of the 'F corpus' (*table D*). Apple and peach

are the most important species both in terms of number of publications and of French production volumes. Publications on table grape benefit from joint studies with wine grape that can be applied to both types of grapes. Although the volume of their production is not significant in France, Citrus and banana are studied mainly in relation to viruses and phytoplasmas.

The thirteen main vegetable species or groups of species studied by INRA (*table II*) represent 80% of the 'V corpus'. Tomato and salads are the most important species both in terms of number of publications and of French production volumes. The large number of publications on eggplant, pepper and sweet pepper is due to the fact that these species are studied for plant-parasite interactions on Solanaceae (Tomato Mosaic Virus and Potato Virus Y).

### 3.3. Main subfields addressed by INRA on fruits and vegetables

Unlike WoS publications, which are assigned to one or more Subject Categories by Thomson Reuters, non-WoS F&V

**Table I.**

Distribution of non-WoS publications of INRA's authors classed by fruit species or group of species and comparison with French production rank (fruit corpus: 1,475 publications, 2002–2006). Rank is based on production in tons in 2006 (SCEES, *Service Central des Études Économiques et Statistiques*, France).

Fruit species	Number of publications <sup>1</sup>	Total fruit publications (%)	Rank / France (only for production > 10 000 t)
Apple	255	17.3	1
Table grape	212	14.4	8
Peach	163	11.1	2
Citrus	143	9.7	11
Apricot	110	7.5	5
Cherry	102	6.9	7
Prune	97	6.6	3
Banana	97	6.6	–
Pear	72	4.9	4
Strawberry	41	2.8	9
Olive	41	2.8	12
Walnut	31	2.1	10
Other species <sup>2</sup>	76	5.1	–

<sup>1</sup> A publication may concern several fruit species.

<sup>2</sup> Chestnut, almond, quince, kiwi, pineapple, fig, mango, etc.

**Table II.**

Distribution of non-WoS publications of INRA's authors classed by vegetable species or group of species and comparison with French production rank (vegetable corpus: 856 publications, 2002–2006). Rank is based on production in tons in 2006 (SCEES, *Service Central des Études Économiques et Statistiques*, France).

Vegetable species	Number of publications <sup>1</sup>	Total vegetable publications (%)	Rank / France (only for production > 100 000 tons)
Tomato	324	37.9	1
Salads <sup>2</sup>	116	13.6	2
Melon	71	8.3	8
Pepper and sweet pepper	65	7.6	–
Eggplant	34	4.0	–
Cabbage	33	3.9	3
Carrot	28	3.3	4
Sweet corn	24	2.8	5
Cucumber	21	2.5	13
Yam	17	1.9	–
Green pea	13	1.5	9
Zucchini	12	1.4	12
Green bean	11	1.3	7
Other species <sup>3</sup>	25	2.9	–

<sup>1</sup> A publication may concern several vegetable species.

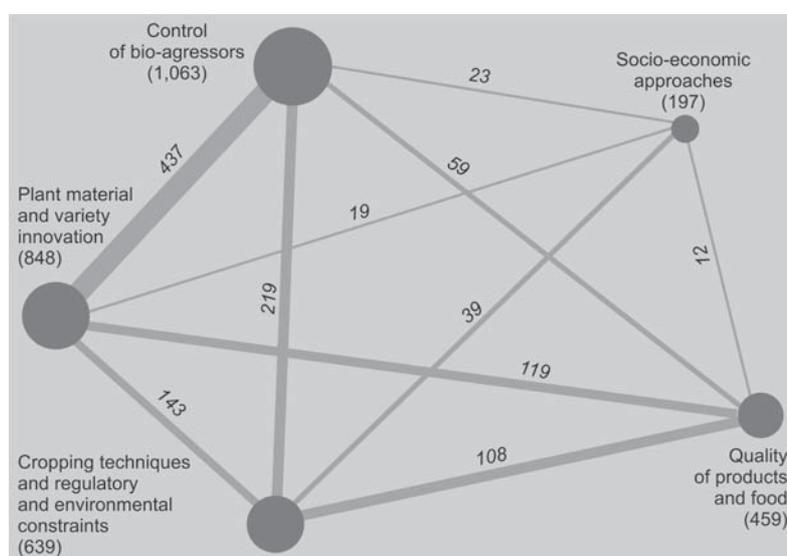
<sup>2</sup> Lettuce, mache, salad, belgian endive.

<sup>3</sup> Leek, garlic, onion, asparagus, etc.

publications are not. To assign non-WoS publications to subfields, we had to carry out a text analysis with the Sphinx Survey software, Lexica option. To delimitate F&V subfields, we used the subfields proposal made by the INRA fruit and vegetable sector group<sup>1</sup>. This step made it possible to classify publications according to five main F&V subfields (*figure 4*).

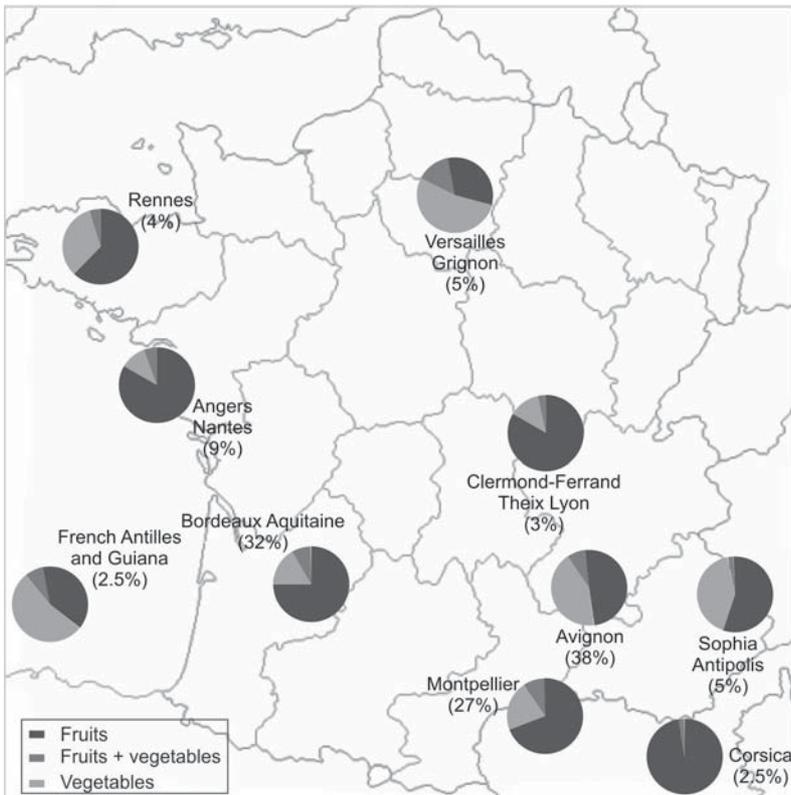
Almost half of the F&V publications focused on “Control of bio-aggressors”, with reference to diagnosis, emerging diseases, pesticide resistance, plant/bio-aggressor interactions, disease resistance and integrated protection.

Research on “Plant material and variety innovation” accounted for 38% of the

**Figure 4.**

Distribution of the non-WoS publications of INRA's authors (France) regarding fruits (F) and vegetables (V) by subject and network of the links between five subjects. Numbers in brackets indicate number of publications for each subject (F&V corpus: 2,224 publications, 2002–2006).

<sup>1</sup> Fiche Fruits et Légumes : les recherches à l'INRA. Les acteurs de l'INRA, INRA, 2007 <http://www.INRA.fr/vegetal/content/download/2919/29334/version/2/file/Fiche-Fruits-legumes.pdf>



**Figure 5.**

Top 10 INRA (France) research centers and distribution of non-WoS publications of INRA's authors classed by fruit (F) and/or vegetable (V) species (F&V corpus: 2,224 publications, 2002–2006).

publications. These publications are linked to genetic resources and selection methods (molecular markers, plant breeding, etc.).

A third of the publications concerned two subfields. Graphic visualization was an original means of representing the strong interaction between “Plant material and variety innovation” and “Control of bio-aggressors” work on one hand, and between “Qualities of products and food” and “Cropping techniques and regulatory and environmental constraints” on the other.

Publications on “Socio-economic approaches” addressed production, markets and trade, fruit and vegetable supply chains, marketing, and quality labels.

### 3.4. INRA fruit and vegetable research network

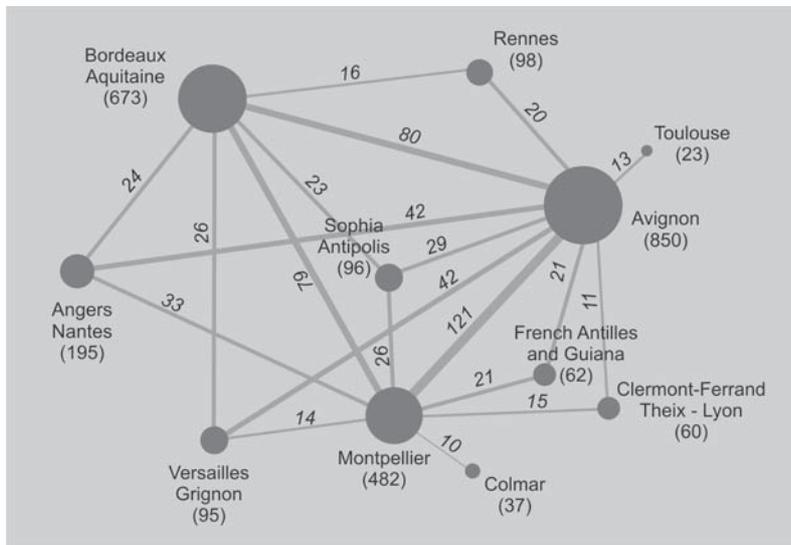
The INRA F&V research network was analyzed from information embedded in the ‘Author Address’ field. Two levels were considered: INRA research divisions and INRA research centers.

#### 3.4.1. INRA research centers

Ten INRA research centers contributed to 90% of the F&V corpus (*figure 5*); the Avignon, Bordeaux Aquitaine and Montpellier research centers published three-quarters of the F&V publications.

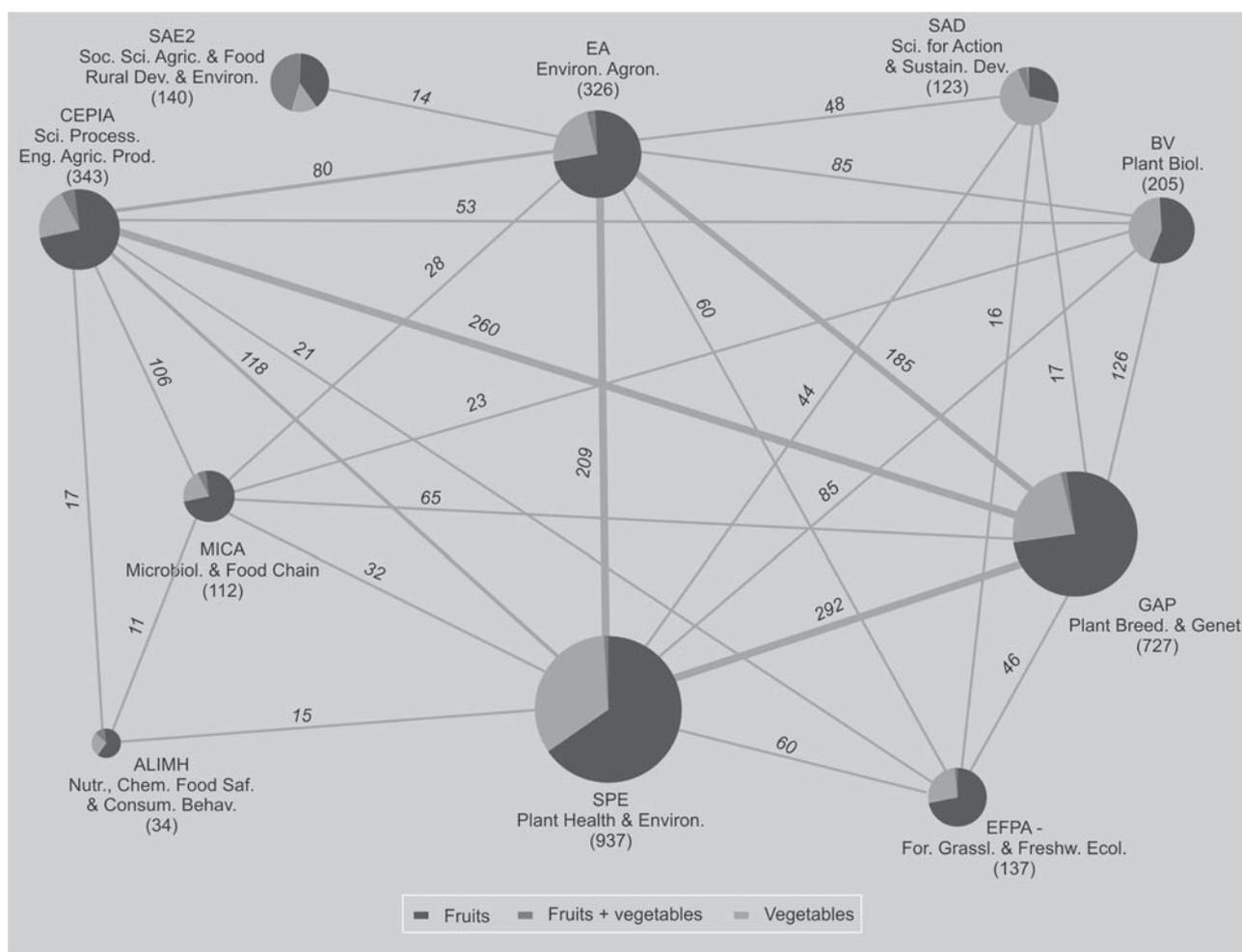
INRA research centers generally work on fruit and vegetable species, with the exception of the Corsica research center, which is specialized in Citrus.

The research centers of Rennes, Versailles Grignon, and French Antilles and Guiana mainly study vegetable species, whereas the research centers of Angers Nantes, Bordeaux Aquitaine and Montpellier mainly study fruit species. The research center in Avignon studies both fruit and vegetable species.



**Figure 6.**

Collaboration network of INRA (France) research centers (number of collaborations > 10). Numbers in brackets indicate number of non-WoS publications regarding fruits (F) and vegetables (V) for each INRA research center (F&V corpus without Theses: 1,757 publications, 2002–2006).



The network of collaborations between INRA research centers shows that three main INRA research centers frequently work together on F&V species and results are disseminated through co-publications (figure 6).

### 3.4.2. INRA research divisions

The distribution of publications by INRA research divisions (figure 7) reveals that the two main research divisions of INRA involved in fruit and vegetable research are Plant Health and Environment (SPE) and Plant Breeding and Genetics (GAP); representing, respectively, 42 % and 33% of the F&V corpus. Environment and Agronomy (EA), Plant Biology (BV), and Science and Process Engineering of Agricultural Products (CEPIA) are less involved in F&V; rep-

resenting, respectively, 22%, 10% and 10% of the corpus.

The interactions between the INRA research divisions (figure 7) point out the importance of multidisciplinary research.

### 3.5. INRA collaborations

We considered a paper to be a co-publication when more than one author's address was provided. For this step, 'Theses' were not included because they are signed by only one author; the F&V corpus therefore was refined to 1,757 publications.

The analysis of authors' addresses allowed the identification of the different partners involved in the publications.

**Figure 7.**

Distribution of fruit and vegetable publications by INRA (France) research division and by fruit (F) and/or vegetable (V) species; collaboration network between INRA research divisions (number of publications > 10). Numbers in brackets indicate number of non-WoS publications for each INRA research division (F&V corpus: 2,224 publications, 2002–2006).

**Table III.**

Distribution of non-WoS publications of INRA's authors regarding fruits (F) and vegetables (V), classed by French partnerships (F&V corpus with a French partner: 503 publications, 2002–2006).

French partner	Number of publications <sup>1</sup>	Total publications co-signed with a French partner <sup>1</sup> (%)	Total F&V publications <sup>1</sup> (%)
Technical centers (Ctifl), experimental stations (CIREF, GRAB, SERFEL, APREL, etc.) and professional bodies (farmers' associations - GRCETA, SRPV, DRAF, etc.)	256	50.9	14.6
Research institutions (Cirad, CNRS, etc.)	181	36.0	10.3
Higher education institutions (universities, Engineers' Higher Education Institutes, etc.)	77	15.3	4.4
Private firms	58	11.5	3.3
Number of publications co-signed with a French partner	503	100	29.0

<sup>1</sup> A publication may be co-signed by several French partners, so the sum of the publications for each group of French partners is superior to the total number of publications.

### 3.5.1. INRA domestic collaborations

Almost 30% of the non-WoS publications were co-signed with a French partner (*i.e.*, 503 publications). The main French INRA partners were technical centers (*e.g.*, Ctifl) and experimental stations (*table III*), with 256 co-signed publications (nearly 15% of the publications). Cirad is INRA's main research institution partner on tropical fruit studies.

### 3.5.2. INRA international collaborations

Nearly a fifth of the F&V corpus (377 publications) was co-signed with an author from a foreign country.

INRA's most important international collaborations are with Italy and Spain (*figure 8*), which are the main foreign partners of INRA and are also the main competitors of France in the fruit and vegetable sector. The USA takes third place due to long-standing partnerships. With the exception of relationships with the Netherlands and Brazil, collaborative work is mainly focused on fruit studies.

## 4. Conclusion

The bibliometric analysis presented highlighted the network of INRA research on

fruits and vegetables and provided complementary information to the previous study of WoS publications [2].

Due to the specific features of the non-WoS publications, additional tools had to be developed to retrieve INRA publications from various sources (institutional repository, lists of publications of the research units), harmonize data, and complete metadata. These preliminary steps allowed us to perform a bibliometric analysis following the same pattern as the F&V WoS publications analysis. Moreover, we combined bibliometric and text analysis with graphic visualizations.

This analysis shows that INRA plays an active role in the fruit and vegetable research field. Research is conducted on the most widely cultivated species in France; for fruit species, these are apple and peach, for vegetable species, tomato and salads.

All of INRA's plant research divisions are involved in F&V research and the research is geographically located in the main research centers of horticultural and/or fruit regions, particularly in South-Eastern France.

The study of INRA's partnerships shows that 29% of the publications were co-signed with a French partner and 21% with a foreign country. This finding reflects the quality and dynamism of the fruit and vegetable



- <http://www.INRA.fr/vegetal/content/download/3136/31408/version/1/file/synthese-biblio-f%26l-2000-2006.pdf>
- [2] Leiser H., Aventurier P., Fournier D., Dosba F., Jeannequin B., Tools for producing indicators from a bibliometric study of scientific production: the case of fruit and vegetable publications by the French National Institute for Agricultural Research (INRA), *Fruits* 64 (2009) 305–312.
- [3] Batagelj V., Mrvar A., Pajek - Program for Large Network Analysis, *Connections* 21 (1998) 47–57.
- [4] Anon., *Bibliometrics as a tool for the analysis of the scientific production of a country*, Obs. Sci. Tech. (OST), Paris, France, 2009, 11 p.
- [5] Jeannequin B., Dosba F., Amiot-Carlin M.J., *Un point sur les filières fruits et légumes : caractéristiques et principaux enjeux*, INRA Ed., Paris, France, 2005.

---

### **Herramientas para analizar y cartografiar las publicaciones científicas no indexadas en la *Web of Science*: ejemplos de publicaciones del instituto nacional francés de investigación agronómica (INRA) sobre frutas y hortalizas.**

**Resumen — Introducción.** El grupo del sector de frutas y hortalizas del INRA realizó, a lo largo del periodo 2002–2006, un estudio bibliométrico de las publicaciones del INRA sobre frutas y hortalizas. Un estudio anterior se había dedicado a las publicaciones indexadas en la *Web of Science*, el trabajo presentado aquí toma en cuenta las publicaciones restantes del INRA, es decir la literatura gris, los artículos en las revistas técnicas y los artículos científicos no indexados en la *Web of Science*. **Material y métodos.** Etapas diferentes permitieron constituir y tratar un corpus de las publicaciones de los investigadores del INRA. El corpus se constituyó consultando el archivo institucional del INRA (ProdINRA). Se llevó a cabo una investigación para recopilar las publicaciones de los investigadores INRA, con el fin de garantizar la exhaustividad de los datos. Dado a los diferentes tipos de publicaciones estudiados, hubo que desarrollar un método original para constituir una base de datos que permitiera un análisis bibliométrico. El programa Sphinx Plus<sup>2</sup> se empleó para el análisis de las publicaciones y el programa Pajek se empleó para visualizar las redes. **Resultados.** Identificamos 2 224 publicaciones INRA en el periodo 2002–2006. Dichas publicaciones se analizaron para poder localizar las especies de frutas y hortalizas estudiadas en el INRA, así como los centros y departamentos de investigación implicados, las temáticas de investigación, y los colaboradores del INRA franceses e internacionales. El uso de representación de redes permitió una visualización de las relaciones entre actores (INRA y fuera del INRA). **Conclusión.** Este trabajo permitió cartografiar las publicaciones del INRA sobre frutas y hortalizas. Se confirmó el papel activo del INRA en la investigación de frutas y hortalizas. En Francia, esta investigación se realiza en los centros INRA en el núcleo de las principales regiones hortícolas y/o fruteras, sobre todo en el sudeste de Francia. A pesar de que las publicaciones estudiadas den relativamente poco lugar a co-publicaciones, este trabajo muestra las relaciones mantenidas por el INRA con otras instituciones francesas, colaboradores del sector y otros países extranjeros.

**Francia / frutas / hortalizas / instituciones de investigación / investigación oficial / publicaciones científicas / colección de datos / ciencia de información / procesamiento de información / coparticipación**

