



HAL
open science

Analyse bibliométrique des publications scientifiques françaises et mondiales sur les protéagineux au cours de la période 2000-2009 : comparaison au soja et aux espèces modèles

Francoise Guillaume, Dominique Millot, Julia Buitink, Jacques Gueguen, Marie-Helene Jeuffroy, Maud Le Gall, Nathalie G. Munier-Jolain, Bernard Tivoli, Gérard Duc

► To cite this version:

Francoise Guillaume, Dominique Millot, Julia Buitink, Jacques Gueguen, Marie-Helene Jeuffroy, et al.. Analyse bibliométrique des publications scientifiques françaises et mondiales sur les protéagineux au cours de la période 2000-2009 : comparaison au soja et aux espèces modèles. *Innovations Agronomiques*, 2010, 11, pp.137-145. hal-01173194

HAL Id: hal-01173194

<https://hal.science/hal-01173194>

Submitted on 30 May 2020

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.



Distributed under a Creative Commons Attribution - NonCommercial - NoDerivatives 4.0 International License

Analyse bibliométrique des publications scientifiques françaises et mondiales sur les protéagineux au cours de la période 2000-2009 : comparaison au soja et aux espèces modèles.

Bridet-Guillaume F.¹, Millot D.², Buitink J.³, Gueguen J.⁴, Jeuffroy M.H.⁵, Le Gall M.⁶, Munier-Jolain N.², Tivoli B.⁷, Duc G.²

¹ : INRA, UAR 116, Equipe Régionale d'Information Scientifique et Technique, F-35000 Rennes, France

² : INRA, UMR 102, Génétique et d'Ecophysiologie des Légumineuses à Graines, BP 86510, F-21000 Dijon, France

³ : INRA, UMR 1191, Physiologie Moléculaire des Semences, 16 Bd Lavoisier, F-49000 Angers, France

⁴ : INRA, UR 1268, Biopolymères Interactions Assemblages, rue de la Géraudière, BP 71627, F- 44000 Nantes, France

⁵ : INRA, UMR 0211, Agronomie INRA-AgroParis Tech, av Lucien Brétignières, F-78000 Thiverval-Grignon, France

⁶ : INRA, UMR 1079, Système d'Elevage, Nutrition Animale et Humaine, Saint-Gilles, F-35000 Rennes, France

⁷ : INRA, UMR 1099, Biologie des Organismes et des Populations appliquée à la Protection des Plantes, BP 35327, F-35000 Rennes, France

Correspondant : duc@dijon.inra.fr

Résumé :

Nous proposons un diagnostic des forces de recherche investies sur les légumineuses protéagineuses importantes pour l'Europe (pois, féveroles, lupins, pois chiches) par une analyse des publications scientifiques au niveau mondial et français sur la période 2000-2009, en regard des investissements sur le soja et sur les espèces modèles. La taille du corpus mondial « protéagineux + soja + haricots + lentilles + *Medicago truncatula* + *Lotus japonicus* + *Arabidopsis thaliana* » a augmenté de 33% entre les périodes 1990-1999 et 2000-2009, avec amplification des contributions sur *Arabidopsis thaliana* et une stabilité dans la répartition entre espèces de légumineuses à graines, qui reste corrélée aux volumes produits. Le corpus mondial identifie sur cette période respectivement 23640, 18946 et 1335 références sur le soja, les protéagineux et les espèces légumineuses modèles. Sur la période 2000-2009 la France contribue à environ 25% des publications mondiales sur les espèces légumineuses modèles (*Medicago truncatula*, *Lotus japonicus*). Environ 10% des publications sur protéagineux font appel à une espèce légumineuse modèle en France à comparer à 1% dans le monde. Les publications sur les espèces légumineuses modèles sont en majorité issues des disciplines de la génétique et de la physiologie végétale. Le pois est l'espèce protéagineuse la plus étudiée en France (génétique, physiologie, protection des plantes, agronomie-écophysiologie sont les disciplines prépondérantes), alors que les travaux sur le soja sont dominants au niveau mondial (les mêmes disciplines sont mobilisées que pour les protéagineux). Une première analyse par discipline et thème permet d'identifier des signaux forts et faibles dans les publications françaises. Nous élargirons prochainement au niveau mondial notre diagnostic par thématique, en le combinant à une étude des blocs géographiques mondiaux impliqués, ainsi qu'à une analyse du réseau des collaborations internationales établies par la recherche française, incluant une description des journaux impliqués.

Mots-clés : légumineuses, protéagineux, protéines, alimentation animale, soja, espèces modèles, *Pisum sativum*, *Vicia faba*, *Lupinus spp*, *Cicer arietinum*, *Glycine max*, *Medicago truncatula*, *Lotus japonicus*, *Arabidopsis thaliana*, publication scientifique, bibliométrie, France, Monde

Abstract: Bibliometric analysis of French and worldwide scientific literature on grain legume crops from 2000 to 2009: comparison to soyabean and model species

We propose a diagnosis of research investments on major grain legumes used for feeds in Europe called grain legumes (pea, faba bean, chickpea) through a bibliometric analysis of scientific papers produced at the world and French levels for the 2000-2009 period, in comparison with studies on soybean and model species. The size of the corpus for "grain legumes + soybean + phaseolus + lens + *Medicago truncatula* + *Lotus japonicus* + *Arabidopsis thaliana*" showed a 33% increase in 2000-2009 in comparison with 1990-1999, with an increased contribution for *Arabidopsis*, and a stable share between grain legume species correlated with production. The world corpus identified 23640, 18946 and 1335 papers on soybean, grain legumes and model legume species respectively. Between 2000-2009, France contributes to ca 25% of the world production on model legume species (*Medicago truncatula*, *Lotus japonicus*). About 10% of papers on grain legumes in France used model legume species compared to 1% at the world level. Papers on model legume species are mostly in plant genetics and plant physiology disciplines. *Pisum* is the major grain legume in French research investments (genetics, physiology, pathology, agronomy-ecophysiology are dominant disciplines) whereas soybean is the main research object at the world level (similar disciplines are used). A preliminary analysis by disciplines and research topics of research has identified strong and weak signals within the French corpus. Our perspective is to extend a similar analysis throughout the different world zones, combining it with a description of the international collaborations developed by French teams and a survey of journals in which papers are published.

Keywords: grain legumes, protein, animal nutrition, soyabean, model species, *Pisum sativum*, *Vicia faba*, *Lupinus spp*, *Cicer arietinum*, *Glycine max*, *Medicago truncatula*, *Lotus japonicus*, *Arabidopsis thaliana*, scientific production, bibliometrics, France, world

Introduction

Le groupe « filière protéagineuse » de l'INRA a conduit récemment une analyse de l'état et des perspectives de la filière (Guéguen et al. 2008) et engage aujourd'hui une évaluation qualitative et quantitative des forces de recherche investies sur les légumineuses protéagineuses importantes pour l'Europe (pois, féveroles lupins, pois chiches). Cette évaluation s'appuie sur une analyse des publications scientifiques de la période 2000-2009 au niveau mondial et français, en regard des investissements sur le soja et sur les espèces modèles. Une étude bibliométrique sur les protéagineux avait été menée sur la période 1990-1999 et constitue un premier référentiel.

1- Constitution des corpus « Monde » et « France »

Nous avons choisi de travailler sur la base CAB Abstract parce qu'elle est plus complète sur notre sujet que le Web of Science (Thomson Reuters ®) et beaucoup plus facile à interroger grâce à son indexation avec des descripteurs (organisés en thésaurus) et des cabicodes thématiques. Le corpus « Monde » est construit à partir de l'interrogation des bases CAB sans exclusion de types de références car les articles de journaux, papiers de conférences, chapitres d'ouvrages et thèses y constituent plus de 95% de la ressource. La construction de nos différents corpus de travail est schématisée dans la Figure 1.

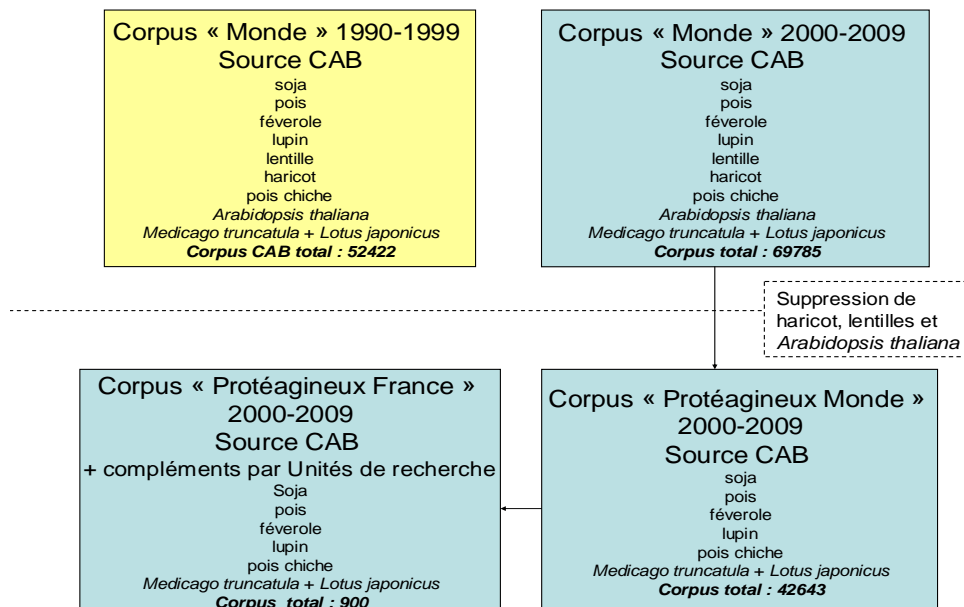
La constitution du corpus CAB « Monde » 2000-2009 a été réalisée à partir d'une interrogation « espèces » sur la base CAB Abstract. Dans un premier temps, notamment pour rester cohérents avec le corpus CAB « Monde » 1990-1999, les espèces : soja (*Glycine max*), pois (*Pisum spp*), féverole (*Vicia faba*), lupin (*Lupinus albus*, *L. luteus*, *L. angustifolius*), lentille (*Lens culinaris*), haricot (*Phaseolus*

vulgaris), pois chiche (*Cicer arietinum*), les légumineuses modèles (*Medicago truncatula*, *Lotus japonicus*) et l'espèce modèle *Arabidopsis thaliana* ont été incluses.

Dans une deuxième étape, pour la constitution du corpus « protéagineux Monde », nous avons exclu du corpus primaire CAB « Monde » 2000-2009 les espèces lentille (*Jens culinaris*), haricot (*Phaseolus vulgaris*) et *Arabidopsis thaliana*.

Pour la constitution du corpus « protéagineux France », nous sommes partis d'une extraction de références du corpus « protéagineux Monde » par sélection positive du mot France dans le champ « adresse ». Mais la base CAB n'attribuant qu'une seule adresse par article (l'adresse de correspondance), nous avons sollicité les différentes équipes de recherches françaises travaillant sur les espèces importantes pour la filière pour compléter ce premier repérage. Nous avons ainsi alimenté notre corpus avec d'autres références des CAB non attribuées à la France et quelques autres non référencées dans les CAB mais considérées comme pertinentes. Ce corpus « protéagineux France » contient 900 références (dont 82% sont signalées dans les CAB). Il a ensuite été enrichi par une indexation manuelle de chaque référence, en termes d'espèces, de disciplines et de thématiques.

Figure 1 : Constitution et dimension des corpus « Monde » et « France »

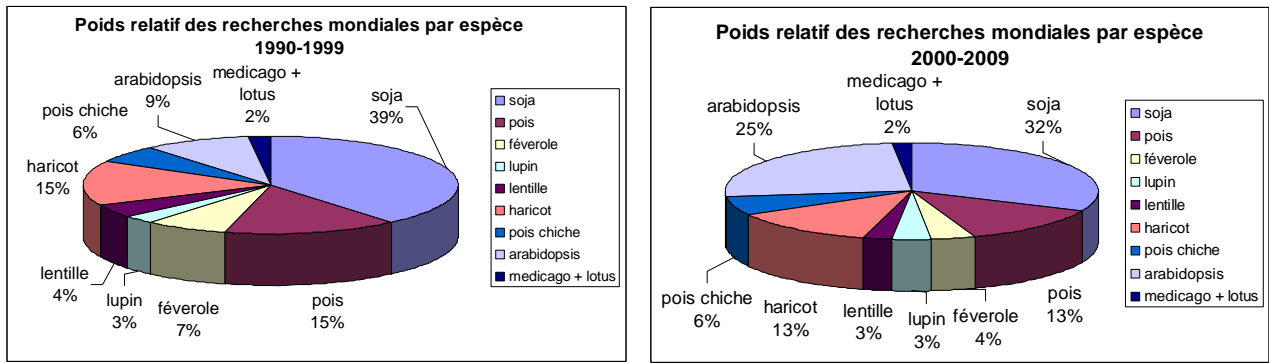


2- Résultats de la répartition en fonction des espèces travaillées

Dans une première étape, deux légumes secs importants en Europe et strictement utilisés en alimentation humaine (haricot et lentille) ont été inclus dans le corpus. Les autres espèces de légumineuses à graine que nous avons considérées sont (1) des légumineuses protéagineuses à usage mixte (alimentation animale et humaine) (pois, féverole, lupins, pois chiches) importantes pour l'Europe, (2) le soja, légumineuse à graine importante dans le monde, (3) les espèces légumineuses modèles (*Medicago truncatula*, *Lotus japonicus*) et (4) l'espèce crucifère modèle *Arabidopsis thaliana*.

Le corpus « Monde » 1990-1999 contient 52 422 références (Guéguen et al., 2000), et celui de 2000-2009 contient 69 785 références, ce qui représente une augmentation de 33% entre les deux périodes étudiées (Figure 2). Celle-ci s'explique par une amplification de 9% à 25% des contributions sur *Arabidopsis thaliana*, espèce modèle importante pour les disciplines de la physiologie et de la génétique végétale.

Figure 2 : Poids relatif des recherches mondiales par espèce pour les périodes 1990-1999 et 2000-2009



Cette première analyse met en évidence un volume important des travaux sur le soja, première légumineuse à graines produite (Tableau 1), oléoprotéagineux très cultivé dans le monde et notamment en zones tropicales. Toujours sur ces deux périodes de référence, on note peu de modifications dans la répartition entre les différentes espèces de légumineuses à graines et des légumineuses modèles. Il y a une très bonne corrélation du nombre des références avec les volumes produits mondialement (cf Tableau 1).

Tableau 1 : Productions mondiales et européennes de légumineuses à graines. (Millions de tonnes, source FAO STAT 2008)

Espèces	Monde	Europe
Soja	230	2,7
Haricot en sec	20,4	0,4
Pois en sec	9,8	3
Pois chiche	8,8	0,06
Fève et féverole en sec	3,7	0,7
Lentille	3,5	0,03
Lupin	1	0,2

Pour la suite de cette analyse, nous distinguerons trois corpus qui se sont révélés de grande taille : les protéagineux (pois, féverole, lupin, pois chiche) (18 946 références), le soja (23 640 références) et les espèces légumineuses modèles *Medicago truncatula* + *Lotus japonicus* (1335 références).

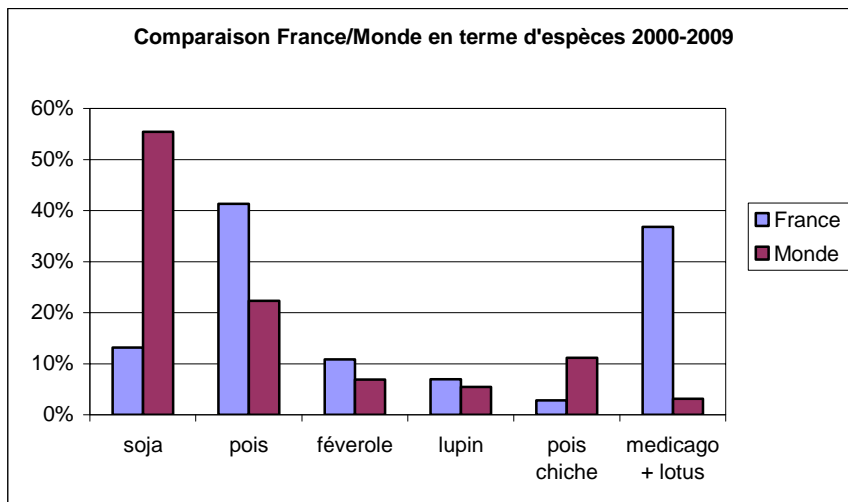


Figure 3 : Comparaison France/ Monde avec un focus sur protéagineux.

Pour cette analyse, 50 références « génériques », non affectées à une espèce en particulier ont été retirées du corpus « protéagineux France », aboutissant à 850 documents.

Des différences méthodologiques entre les analyses bibliométriques conduites sur les deux périodes ne

permettent pas des comparaisons complètement rigoureuses. Néanmoins, une première estimation indique que les tailles des corpus « protéagineux France » entre les deux décennies sont du même ordre de grandeur.

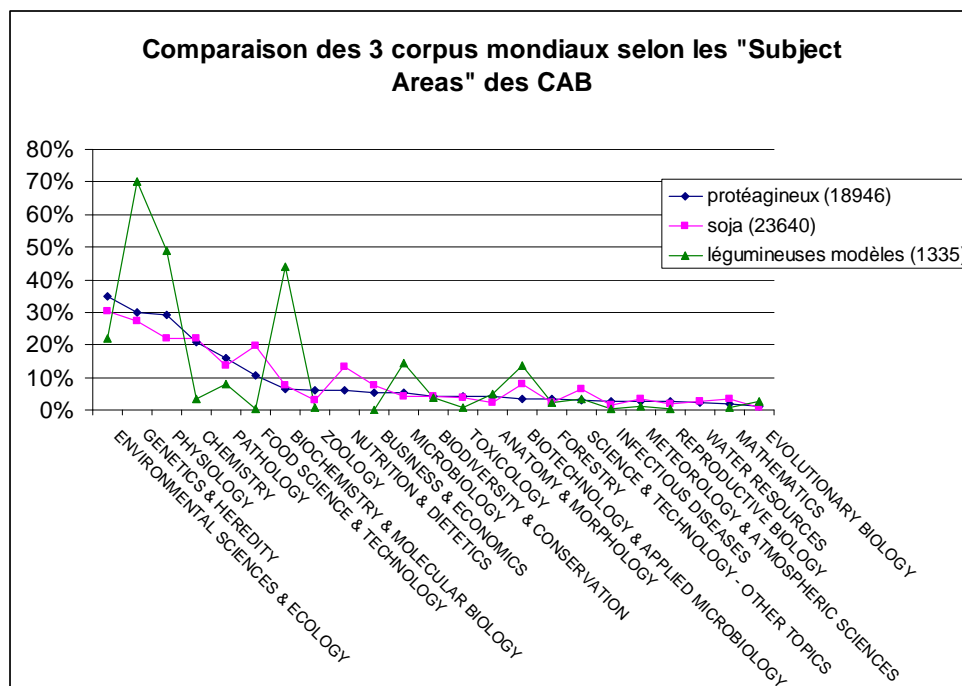
Par rapport au Monde, on mesure sur le corpus « protéagineux France » (Figure 3) une moindre proportion de références sur le soja et les contributions scientifiques sur les différentes espèces de légumineuses à graines sont corrélées aux surfaces cultivées en France. La prépondérance du pois est très visible.

On note aussi une forte production scientifique française sur les espèces légumineuses modèles (*Medicago truncatula*, *Lotus japonicus*). Sur la période 2000-2009, la France a contribué à environ 25% des publications mondiales sur ces espèces.

3- Résultats de la répartition en fonction des disciplines étudiées

L'allocation par discipline et thème du grand corpus « protéagineux Monde » est basée sur une indexation par le thesaurus des CAB, alors que chaque publication de la base « protéagineux France » a fait l'objet d'une révision et d'un ajustement de l'indexation CAB par une attribution « manuelle » de disciplines.

Figure 4 : Comparaison des trois corpus mondiaux « Soja », « Espèces protéagineuses » et « Espèces modèles » par l'analyse des « Subject Areas » (disciplines) des CAB



Sur le corpus « protéagineux Monde », on observe (Figure 4) une répartition assez voisine des disciplines entre le soja et les protéagineux. Les « sciences de l'environnement », la « génétique », « physiologie », « biochimie-biologie moléculaire », « pathologie » sont prépondérantes. Cependant, pour le soja « alimentation-technologie » et « nutrition-diététique » sont plus importantes que sur les protéagineux. Sur les espèces légumineuses modèles, « génétique », « physiologie », « biochimie-biologie moléculaire », « environnement-écologie » sont les disciplines les plus appelées.

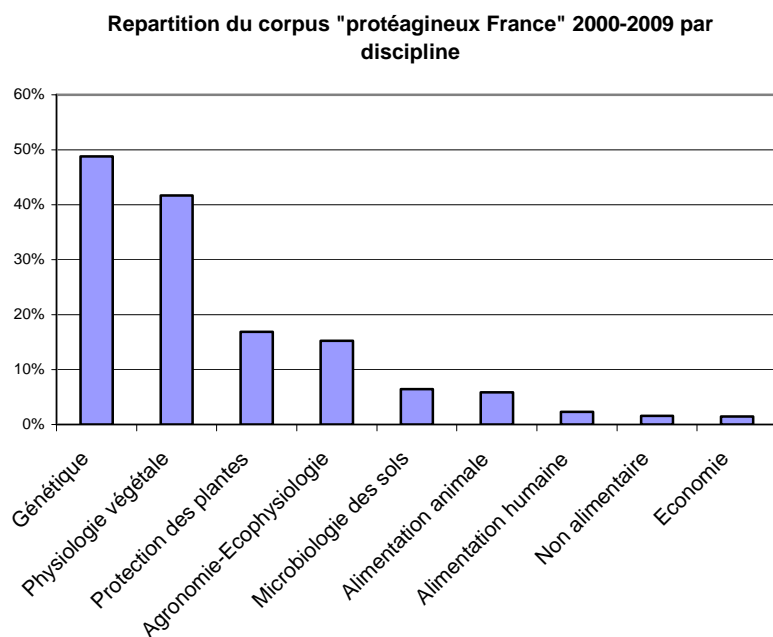


Figure 5 : Répartition du corpus « protéagineux France » 2000-2009, en fonction des disciplines attribuées « manuellement » aux références du corpus

Les disciplines prépondérantes en France (Figure 5) sont les mêmes que dans le Monde (génétique, physiologie et protection végétales, agronomie-écophysiologie, microbiologie des sols). Les développements de la génétique moléculaire, avec des possibilités grandissantes dans les études d'expression des gènes, ont rapproché les disciplines de la génétique et de la physiologie. On note peu d'investissements en matière de modes d'utilisation et transformation, ainsi que dans les sciences économiques.

4- La mobilisation des espèces modèles dans les recherches sur protéagineux

Pour cette analyse, 50 références « génériques », non affectées à une espèce en particulier ont été retirées du corpus « protéagineux France », aboutissant à 850 documents.

Tableau 2 : Comparaison Monde/France pour la mobilisation des espèces modèles, par les travaux sur protéagineux ou soja (nombre de références)

MONDE	Protéagineux	Soja	Dont travaux utilisant : <i>M. truncatula</i> et <i>L. japonicus</i>	Dont travaux utilisant : <i>A. thaliana</i>
Protéagineux	18946	1097	106	245
Soja		23640	92	209
<i>M. truncatula</i> et <i>L. japonicus</i>			1335	127
FRANCE	Protéagineux	Soja	Dont travaux utilisant : <i>M. truncatula</i> et <i>L. japonicus</i>	Dont travaux utilisant : <i>A. thaliana</i>
Protéagineux	457	27	49	26
Soja		112	9	8
<i>M. truncatula</i> et <i>L. japonicus</i>			372	25

La différence d'indexation des « protéagineux Monde » et « protéagineux France » génère un léger biais dans leur comparaison. Sur la période 2000-2009, le Tableau 2 montre que la France contribue à environ 25% des publications mondiales sur les espèces légumineuses modèles (*Medicago truncatula*, *Lotus japonicus*). Environ 10% des publications sur les protéagineux font appel à une espèce légumineuse modèle en France, contre seulement 1% au niveau mondial. Ce résultat est révélateur d'une stratégie méthodologique choisie en France, qui est d'exploiter la plante modèle pour des approfondissements dans différentes disciplines et de s'appuyer sur les fortes homologues de séquences et synténies entre espèces cultivées et espèces modèles pour faciliter la recherche des gènes.

5- Analyse des thématiques de recherche dans le corpus « protéagineux France » 2000-2009

On rappelle que chaque publication du corpus « protéagineux France » a fait l'objet d'une indexation par une attribution « manuelle » de thématiques de recherche. Des thématiques ont été ainsi classées par nombre décroissant de références, dans les principales disciplines (Tableau 3).

Tableau 3 : Les 6 thèmes de recherche les plus appelés pour les disciplines les plus représentées. (effectifs d'occurrences dans chaque discipline, dans le corpus « protéagineux France » 2000-2009 de 900 références)

Disciplines	Génétique végétale plantes (439)	Physiologie végétale (375)	Protection des végétaux (152)	Agronomie écopysiologie (137)	Microbiologie des sols (58)
Thèmes	Symbioses et acquisition des ressources (180)	Symbioses et acquisition des ressources (214)	Tolérance aux stress biotiques-abiotiques (139)	Symbioses et acquisition des ressources (43)	Symbiose et acquisition des ressources (53)
	Outils de biologie moléculaire (83)	Tolérance aux stress biotiques-abiotiques(64)	Protection des insectes (69)	Tolérance aux stress biotiques-abiotiques (36)	Biodiversité microflore (34)
	Tolérance aux stress biotiques-abiotiques (78)	Architecture-Phénologie (40)	Protection des champignons (52)	Fertilité des sols (33)	Tolérance aux stress biotiques-abiotiques (9)
	Architecture-Phénologie (51)	Germination (38)	Protection des adventices (19)	Systèmes de culture (28)	Fertilité des sols (7)
	Composition de la graine (49)	Composition de la graine (33)	Biodiversité des agresseurs (14)	Architecture (24)	Protection des Champignons (4)
	Diversité-évolution (41)	Photosynthèse (32)	Systèmes de culture (10)		Systèmes de culture (2)

Le thème « symbioses et acquisition des ressources » est très dominant, transversal aux disciplines et est surtout représenté par des travaux qui portent sur l'acquisition des ressources azotées. Le thème de la « tolérance aux stress biotiques et abiotiques » est également très représenté dans les disciplines dominantes.

Soulignons tout d'abord que ce résultat est fortement inféodé au mode d'indexation que nous avons constitué et sur lequel il s'appuie. Un développement important sur un thème, dépendra du champ plus ou moins large couvert par celui-ci. Il pourra aussi refléter une réponse forte à une demande forte de la filière sur un thème crucial, ainsi que le développement d'une spécialité reconnue à une équipe de recherche particulière.

Tableau 4 : Les signaux faibles ou forts parmi les thématiques du corpus. (effectifs d'occurrences dans chaque discipline, dans le corpus « protéagineux France » 2000-2009 de 900 références)

Corpus « protéagineux France » (900)	Thèmes
Signaux forts	<ul style="list-style-type: none"> - Symbioses et acquisition des ressources (311) - Tolérance aux stress biotiques-abiotiques (245) - Composition de la graine (118)
Signaux faibles	<ul style="list-style-type: none"> - Impacts environnementaux (30) - Transformation technologique (28) - Utilisation de l'eau (22) - Protection des adventices (20) - Nutrition et santé (17) - Economie-Filière (13) - Associations (6) - Pollinisation (3) - Agriculture biologique (3)

Dans le corpus « protéagineux France », le thème « tolérance aux stress biotiques et abiotiques » fait l'objet de nombreux travaux, certainement en réponse aux problèmes d'instabilité des rendements rencontrés par les cultures de protéagineux. Le thème « impacts environnementaux » est peu appelé directement, mais est néanmoins fortement travaillé dans les nombreuses études qui concernent les symbioses et l'acquisition des ressources. L'absence de fertilisation azotée des cultures de légumineuses permise par la fixation symbiotique de l'azote atmosphérique a en effet des conséquences environnementales très positives (cf autres présentations dans ce numéro). On doit cependant souligner que d'autres thématiques relatives à l'analyse des impacts environnementaux sont faiblement appelées par les travaux inventoriés, telles que l'utilisation de l'eau, la protection des adventices, les associations, la pollinisation, l'agriculture biologique. Le thème « composition de la graine » fait l'objet de nombreux travaux mais en aval, les recherches sur la valeur d'utilisation et les technologies de transformation, sont quant à elles encore peu couvertes. Les analyses de la filière par les sciences économiques et sociales sont elles aussi peu nombreuses.

Conclusions et perspectives

Soulignons tout d'abord que des délais de publication d'un travail scientifique sont longs à partir de son initiation. Trois ans est un minimum, et 5-7 ans sont fréquents dans certaines disciplines. Ainsi, les corpus que nous avons étudiés reflètent davantage des investissements de recherche 1995-2005 et n'incorporent pas d'éventuelles inflexions récemment engagées.

Notre analyse indique que la France a maintenu un investissement sur les protéagineux qui paraît assez stable au plan quantitatif depuis les années 1990, et la répartition entre espèces protéagineuses est proportionnelle aux surfaces cultivées. Par rapport à la période 1990-1999, on note pour le pois et les espèces légumineuses modèles une augmentation dans les disciplines prépondérantes. Cette évolution est très visible au plan international puisque la France contribue à environ 25 % des publications mondiales sur légumineuses modèles.

Dans les publications françaises, une première analyse par discipline et thème permet d'identifier des signaux forts pour les actions qui visent la stabilité de la productivité des protéagineux. Certains thèmes importants pour l'évaluation et la maîtrise des impacts environnementaux, ou pour de nouvelles utilisations conférant de la valeur ajoutée, sont par contre encore peu travaillés.

Nous élargirons prochainement au niveau mondial cette analyse par thématique, en la combinant à une étude des blocs géographiques mondiaux impliqués, et nous décrirons le réseau des collaborations internationales établies par la recherche française. Nous analyserons également les types de journaux supports de ces publications et leurs impacts.

Remerciements

Nous remercions Martine Georget animatrice des groupes filières INRA, Odile Hologne et Dominique L'Hostis de la DV-IST, pour leur appui et leurs encouragements dans ce travail bibliométrique.

Références bibliographiques

FAO Statistiques agricoles 2008. <http://faostat/fao.org/>

Guéguen J. et al., 2000, - La recherche dans la filière protéagineux, document Filière IINRA, 165pp.

Guéguen J., Duc G. (coordonnateurs), Boutin J.P., Dronne Y., Munier-Jolain N., Sève B., Tivoli B., 2008. La filière protéagineuse. Quels défis ? Editions Quae, pp. 147p. *Éditions Quae ISBN 13 978-2-7592-0072-6.*