

Genetic patents in plant biotechnology

Koen Jonkers^{*1}, Catalina Martinez^{**}

*koen.jonkers@ec.europa.eu

European Commission, Joint Research Centre, Institute for Prospective Technological Studies (IPTS),
C/ Inca Garcilaso No. 3, E-41092 Seville, Spain

** catalina.martinez@csic.es

Institute of Public Goods and Policies, CSIC-IPP, C/ Albasanz 26-28, Madrid, 28037, Spain

Introduction

Genetic patenting is an area of heated debate in scientific, legal and economic fora, especially as regards the human genome, but increasingly also with respect to plant biotechnology given the large economic stakes involved (Louwaars et al, 2009; Baillie & Connett Porceddu, 2012). In 2008, the global seed industry had an annual turnover of 2.7 billion Euro. These seeds do not only form the basis for a total product market with an annual turnover of 250 billion Euro, but are of central importance for the global food supply (Louwaars et al, 2009). Intellectual property rights on plants are governed through two different regimes: plant breeders' rights and patents. This paper focuses on the latter.

Our aim is to assess whether policy relevant questions in the field of plant biotechnology, such as ownership of genetic patents, science-industry links and the relation between patents and follow-on research, can be addressed with publicly available data in patent databases linked to DNA sequence repositories. In doing so, we first present a review of the literature and background on the relation between plant biotechnology and genetic patenting, and then we examine the distribution of patent applications claiming plant gene sequences by major patent office and type of organism, number of sequences per patent, type and country of origin of applicants as well as other patent features. We pay special attention to the distinction between patents including genes of the model organism *Arabidopsis thaliana* – as a model organism it is a research tool and therefore a potential indicator for science relatedness – and patents including genes of commercial crops, as well as to the dynamics of public/private nature of ownership and business concentration in the plant seed sector, where five companies hold 64% of all plant sequence EPO applications and over 90% of all the plant gene sequences listed in them. We conclude with a more in-depth exploratory quantitative and qualitative analysis of the value and scope of protection of EPO filings listing gene sequences of different kinds of plant organisms.

Database construction and preliminary findings

In order to build our database, we first downloaded all the patent sequences files (GBPAT) from NCBI GENBANK flat file release 183.0 (February 15 2011 and April 15 2011)² which were used to create a relational database and linked those files to patent information from

¹ The information and views set out in this publication do not necessarily reflect the opinion of the European Commission. The EC does not guarantee the accuracy of the data included in this study. Neither the EC nor any person acting on its behalf may be held responsible for the use which may be made of the information contained herein.

² <ftp://ftp.ncbi.nih.gov>

PATSTAT (October 2010), using patent numbers. GENBANK has 19,189,921 different gene sequence entries and 424,238 patent publication numbers from different patent offices. For other studies using GENBANK see e.g.: Arnaud-Haoud *et al.* 2011 on marine species. The data contained in NCBI GENBANK allows us to come to an alternative, more specific, delimitation of life science patents than would be possible on the basis of IPC classes. Only patents that contain genetic sequences are studied in this paper.

A first noticeable finding is that GENBANK does not have information on USPTO gene sequences broken down by type of organism.³ We therefore focus our analysis on plant sequence patents filed at EPO and through the PCT route worldwide (table 1). In EPO patent applications, plant sequences comprise around 4.1% of the total number of patented sequences, in PCT filings this share is around 3.3 %. This makes plantae one of the largest source groupings after bacteria, mammalia, and synthetic constructs.

Table 1. Patents containing gene sequences and number of sequences by type of filing and source organism.

	EPO filing		PCT filing	
	number of sequences	number of patents	number of sequences	number of patents
Unclassified	47664	1823	81905	5540
Synthetic	570022	6969	2074033	21481
Mammalia	2231844	3316	2186310	13469
Bacteria	227932	1352	146193	3929
Plantae	133995	581	175420	2376
Virus	9849	562	71685	1484
Fungi	18344	385	431636	1457
Arthropoda	10201	126	43399	420
Archae	4590	94	1886	258
Other	8530	532	12868	1349
Total	3262971	15740	5225335	51763

Table 2 lists the top ten source plants, accounting for around 81% of all plant sequences in EPO patent applications and 77% in PCT patent applications. In terms of gene sequences, Arabidopsis is the most important source organism in PCT filings with plant gene sequences, with 20% of all, and the second most important at EPO, with 18%, after Maize with 28%. In terms of patent applications, Arabidopsis ranks first, as 35% of all EPO applications and 39% of all PCT filings include at least one sequence of Arabidopsis.

³ We explored the use of BLAST searches to classify a subsample of the unidentified sequences by type of organism, as a pilot exercise. This proved to be quite a labour intensive effort and will be described in detail in the paper.

Table 2. Top 10 plant gene sequences in patent filings, by organism.

	EPO sequences		EPO patents		PCT sequences		PCT patents	
	Count	% Total Plantae	Count	% Total Plantae	Count	% Total Plantae	Count	% Total Plantae
Maize	37293	27.8%	123	21.2%	28056	16.0%	551	23.2%
Arabidopsis	23845	17.8%	204	35.1%	34613	19.7%	917	38.6%
Soybean	19644	14.7%	76	13.1%	12863	7.3%	331	13.9%
Rice	19614	14.6%	126	21.7%	30793	17.6%	510	21.5%
Wheat	2338	1.7%	83	14.3%	5470	3.1%	263	11.1%
Tomato	1869	1.4%	60	10.3%	2569	1.5%	220	9.3%
Tobacco	1748	1.3%	66	11.4%	6625	3.8%	242	10.2%
Potato	1108	0.8%	58	10.0%	1471	0.8%	210	8.8%
Barley	968	0.7%	51	8.8%	12687	7.2%	176	7.4%
Pea	494	0.4%	23	4.0%	498	0.3%	72	3.0%
Top 10 / Total Plantae		81%				77%		
Total Plantae	133995	100%	581	100%	175420	100.0%	2376	100%

Taking the 581 EPO filings including plant gene sequences shown in Table 2, we perform a more in-depth quantitative and qualitative analysis of different patent features. Several preliminary findings are worth mentioning:

- First, patent applications filed by public research organisation (PRO), alone or jointly with firms, have grown in recent years, although companies are the main applicants of plant genetic sequence patents in EPO.
- Second, small firms and PROs are significantly more likely to take the PCT route than large firms, which reflects the role of the PCT route as a way to gain time and look for funding to proceed with the patenting process and commercialisation of the protected invention, which is especially relevant for financially constrained small firms and PROs.
- Third, a handful of companies own the majority of the patents, following an intense process of concentration in which large agrochemical and pharmaceutical firms have bought up many smaller seed companies and biotech firms. We trace this process and find that at present the top five companies (Monsanto, Bayer, Basf, Syngenta and Dupont Pioneer) and their subsidiaries (including firms that have been bought by these five firms or are under shared ownerships) control 64% of all plant sequence patents in EPO corresponding to over 90% of the sequences for which a patent application has been made at EPO.
- Fourth, preliminary results of econometric estimations on the relation between patent characteristics and number of forward citations received show that patents filed by Bayer and Syngenta are more likely to receive a higher number of forward citations than those filed by BASF, conditional on other patent characteristics. Based on a 2009 report by Louwaars et al, we believe these firms have different business models: Syngenta and Bayer Cropscience develop seeds whereas BASF is into patenting traits and licensing these patents to seed companies. These different business models are also visible in simple correlation analysis, where we observe that BASF patent filings also tend to have a larger number of gene sequences than those of the other top companies.

- Fifth, Arabidopsis gene sequences, which are included in 35% of all the patents in our sample, seem to have a negative effect on the number of forward patent citations received by the patents in our sample. This may be due to its use as research tool in a broad range of applications or the lack of direct commercial relevance of this organism.

Further quantitative and qualitative analyses, including an assessment of the specific subject matter and scope of the EPO filings included in the sample, are currently being undertaken to better interpret these preliminary results and the forces at play in the field of plant biotechnology in relation to genetic patenting.

Acknowledgements

We thank Stéphane Maraut for providing IT help and advice throughout the project. Laura Manzano Baena and Maria Eugenia Espinosa Calvo helped with the codification of USPTO patents. We also wish to acknowledge funding from CSO2011-2350, CSO2009-10845 and CSO2012-32844.

References

Arnaud-Haond, S., Arrieta, JM, Duarte, C.M., 2011, Global Genetic Resources - Marine Biodiversity and Gene Patents, *Science*, Vol. 331 no. 6024 pp. 1521-1522

Baillie, B., Connett Porceddu, M., 2012, Patent Landscape of the Arabidopsis genome and related plant species, <http://www.patentlens.net/daisy/ArabidopsisGenome/1662.html>

Louwaars, N., Dons, H., van Overwalle, G., Raven, H., Arundel, A., Eaton, D., Nelis, A., Breeding business, the future of plant breeding in the light of developments in patent rights and plant breeder's rights, CGN Report 2009-14, Centre for Genetic Resources, the Netherlands.

NCBI, 2012, BLAST - Basic Local Alignment Search Tool, available at: <http://blast.ncbi.nlm.nih.gov/>