

Science-Matrix

Scientometric Study

September 2004



Trends in Forest Genomics in Canada and Other Leading Countries Overall and by Specialties – 1992-2003

Prepared for the
Canadian Forest Service
Natural Resources Canada

Science-Metrix

Scientometric Study

Trends in Forest Genomics in Canada and Other Leading Countries Overall and by Specialties – 1992-2003

September 2004

David Campbell and
Grégoire Côté

Prepared for the
Canadian Forest Service
Natural Resources Canada



Science-Metrix specializes in the measurement and evaluation of science, technology and innovation. Our data collection and assessment methods include bibliometrics, scientometrics, technometrics, surveys and interviews, environmental scans, monitoring and intelligence gathering. We perform program and policy evaluations, benchmarking and sector analyses, market studies and strategic planning. Science-Metrix has a robust knowledge of life and environmental sciences.

514.495.6505 • 4572 avenue de Lorimier • Montréal • Québec • Canada • H2H 2B5
info@science-metrix.com • www.science-metrix.com

Executive Summary

This study measures the scientific output of Canada in the field of forest genomics compared to that of other leading countries. The statistics presented here are based on a count of scientific papers, whose titles contain forest- and genomics-specific keywords. The information was retrieved from the Thomson ISI's Science Citation Index Expanded database, which indexes papers from approximately 5,900 of the world's leading scholarly science and technical journals, covering more than 150 disciplines. These journals reflect significant scientific achievements and are the most widely cited journals in their respective disciplines, comprising more than 80% of the world's citations. Particular attention was paid to ensure unbiased coverage for each country in the selection of keywords. Thus, the resulting dataset does not list every paper for every country; rather, it provides an unbiased comparison of the production of countries active in forest genomics.

The study, which covers a 12-year period (1992–2003), charts the evolution of scientific output in the field at the world level and in leading countries. The evaluation of the level of scientific performance is calculated on the basis of four scientometric indicators (number of papers, papers per capita, specialization index and average of relative citations) and a multicriteria ranking.

Canada, whose commercial forest ranks third in terms of worldwide importance and covers about 235 million hectares, ranked 1st in the first half of the period studied (1992–1997). Subsequently, Canada slipped into 2nd place in the 1998–2000 period, on an even level with France, and to 4th place in the 2001–2003 period, behind France (1st), Sweden (2nd) and Australia (3rd). Overall Canada ranks 2nd for the 12-year period, behind France, which started out in 5th place and finished 1st. When dissecting the field of forest genomics, Canada excels in tree genomics and genomic studies of parasitic tree fungi, whereas it lags behind other countries in genomic studies of parasitic bacteria of trees or tree pests. Canada also has potential strengths in genomic studies of tree pests (insects) and viruses of trees or tree pests. Overall, Canada ranked high in terms of its number of papers (2nd), number of papers per capita (3rd) and specialization index (1st). Therefore, the factor that accounts for Canada's 2nd place in the multicriteria benchmarking is the relatively small citation score of its papers (11th). Consequently, Canada could improve its position and the reach of its studies by maintaining the current level of funding, which is substantially higher than when Canada embarked on its downward trend in terms of scientific impact. This will ensure that Canadian researchers continue to have access to the modern equipment that is essential when pursuing research in a field that is undergoing rapid technological development.

Acknowledgements

This research has benefited from the financial support of Genome Canada for the maintenance of Science-Metrix' biotechnology and genomics database, a subset of the Science Citation Index Expanded database which is used under license from Thomson-ISI.

Key Findings

Global trends in forest genomics

- The number of scientific papers published in genomics as a whole increased rapidly at the world level from about 39,000 papers in 1992 to almost 60,000 papers in 2003, an increase of nearly 60%.
- Within genomics, forest genomics is a small niche with a scientific production accounting for about 1% of the worldwide output in genomics overall in 2003.
- Although more uneven than in genomics, the growth of scientific output in forest genomics over the observed time-frame has also been substantial, with the number of papers per year increasing from 257 in 1992 to 542 in 2003, an increase of 110%.

Benchmarking leading countries in forest genomics

- The United States (US) dominates forest genomics in absolute output, holding about 30% of the world's scientific production with 1,372 papers published over the 12-year period studied.
- With 624 papers, about half as much as the US' output, Canada ranks 2nd overall in the period spanning 1992 to 2003. Japan and France follow closely with respectively 529 and 515 papers.
- Among less productive countries, Brazil and India significantly increased their scientific output during the 2001–2003 period, both with an average annual growth of 24%.
- When considering each country's population size, larger countries with high absolute production such as the US, Japan, France, the UK (UK) and Germany are being overtaken by smaller countries with low absolute scientific production. Sweden and Finland, who rank 8th and 13th in absolute output, rank 1st and 2nd in output per capita.
- Canada and Australia occupy the 3rd and 4th place in relative output.
- Over the 12-year period, Canada (1st), Finland (2nd), Sweden (3rd) and Australia (4th) were the countries with the strongest emphasis on forest genomics being at least twice as specialized as the world average in the field.
- The US, Germany and the UK are the least specialized of the leading countries, being at least 30% less specialized than the world average.
- With respect to average of relative citations, Spain, which already made good progress with respect to the previous indicators, started in 14th place in the 1992–1994 period and gained 12 positions to finish 2nd behind France in the 2001–2003 period. Over the 12-year period, France ranks 1st and Spain 2nd.
- Canada which usually performs very well in terms of average of relative citations ranked 11th with its papers in forest genomics being less cited than the world average over the 12-year period studied.
- France and Canada undoubtedly played a major role in the development of the field, ranking 1st and 2nd in the multicriteria analysis for the 12-year period.
- Globally, Canada lost some ground over time. Starting 1st in the 1992–1994 and 1995–1997 periods, it subsequently slipped to 2nd place in the 1998–2000 period, on a par with France, and to 4th in the last period, behind France (1st), Sweden(2nd) and Australia (3rd).
- In the benchmarking analysis, Sweden and Australia shared 3rd place, the US is 5th, Finland 6th, Japan 7th, the UK 8th, Spain 9th, Germany and Italy both 10th, the Republic of Korea 12th, China 13th, and Brazil is ex aequo with India in 14th position.
- The progress made by Spain has been tremendous. The country improved its position in forest genomics with respect to each indicator. According to the multicriteria ranking, Spain started last in the 1992–1994 period and finished in 6th place in the last period, for an overall rank of 9.

Analysis of strengths and weaknesses

- Canada's main strength lies in the fungus genomics specialty (59 papers, 2nd rank), which involves studying the genetics of tree diseases caused by fungi.
- Although Canadian research in tree genomics, as inferred from ARC, has the lowest impact among the top eight leaders and ranks slightly below the world average, the country is nonetheless more specialized in this field than the world average and ranks 2nd in terms of the number of papers produced in this specialty (443 papers). Thus, tree genomics can be considered as an important strength for Canada.
- Canada has potential strengths in the insect genomics and virus genomics specialties. In insect genomics (40 papers, 3rd rank), Canada's scientific impact (ARC) is more than 1.5 times higher than the world average (the highest among specialties for Canada), but it is not specialized in this field ($SI < 1$), meaning that it does not produce as many papers as would be expected given its share of the world scientific production overall. In virus genomics (47 papers, 2nd rank), the picture is reversed, with Canada's proportion of papers in the field being 1.5 times higher than the world's and its ARC being less than 1.5 times that of the world.
- Canada's main weakness lies in bacteria genomics (17 papers, 7th rank).

Table of contents

Executive Summary	i
Acknowledgements	i
Key Findings	ii
Table of contents	iv
Tables	iv
Figures	iv
1 Introduction	1
2 Global trends in forest genomics	4
3 Benchmarking leading countries in forest genomics	6
3.1 Number of papers in forest genomics	6
3.2 Number of papers in forest genomics per capita	8
3.3 Specialization index	9
3.4 Average of relative citations	11
3.5 Multicriteria analysis	12
4 Analysis of Strengths and Weaknesses	14
5 Conclusion	18
References	19
Appendix: Methods Used in Scientometric Analyses	20

Tables

Table I	Number of papers in forest genomics by leading countries per three-year period, 1992–2003	7
Table II	Ranking of leading countries according to their number of papers in forest genomics, 1992–2003	7
Table III	Number of papers in forest genomics by leading countries per ten million inhabitants per year (annual average of periods), 1992–2003	8
Table IV	Ranking of countries according to number of papers in forest genomics per million inhabitants (annual average of periods), 1992–2003	9
Table V	Leading countries' specialization index in forest genomics per three-year period, 1992–2003	10
Table VI	Ranking of leading countries according to their specialization index in forest genomics, 1992–2003	10
Table VII	Leading countries' average of relative citations in forest genomics per three-year period, 1992–2003	11
Table VIII	Ranking of leading countries according to their average of relative citations in forest genomics, 1992–2003	12
Table IX	Multicriteria ranking of leading countries in forest genomics per three-year period, 1992–2003	13
Table X	Number of world papers in forest genomics by specialty, 1992–2003	15

Figures

Figure 1	Papers in genomics in SCI Expanded, 1992–2003	4
Figure 2	Papers in forest genomics in SCI Expanded, 1992–2003	5
Figure 3	Top 30 countries by number of papers in forest genomics, 1992–2003	5
Figure 4	Leading countries' strengths and weaknesses in forest genomics, 1992–2003	17

1 Introduction

Forests and trees have always been and still are of major importance to humans, providing food, fuel wood, building materials, paper and other products. Forests are also of unique ecological value, providing shelter for most of the planet's biological diversity and playing an essential role in the carbon cycle. However, with the growth of human population, the ever increasing demand for wood products constitutes a major threat to native forests worldwide. Currently, there is an estimated 3.4 billion hectares of forest cover on earth (Sedjo and Botkin, 1997), of which about 14 million hectares are destroyed each year in developing countries (Marcoux, 2000). In Canada alone, 18.6 million hectares were affected by insects in 2001, and 2.8 million hectares were lost due to fire in 2002¹. Adverse effects of this deforestation comprise the emission of about 1.7 Gt of carbon annually (20% to 30% of annual carbon flux) and the extinction of between 14,000 and 40,000 species each year (Kremen *et al*, 2000). With increased public awareness of such problems, the need to rely on alternative sources to old-growth forests and secondary-growth forests is now amply recognized.

Perhaps the most promising alternative to relieve pressure on native forests and to create natural reserves for conservation lies in highly managed tree plantations. It is estimated that only 5% to 10% of the world's forest area would be sufficient to sustainably satisfy world demand using highly managed tree plantations compared to the current model which requires that we exploit 20% to 40% of the world's natural forests (Sedjo and Botkin, 1997). Moreover, plantation forestry provides many economic advantages over traditional harvesting. For one thing, plantations are more accessible and by concentrating resources in a small area, the need for complex road networks is limited and, consequently, transportation costs are significantly reduced. Plantation forestry also allows the selection of fast-growing tree species adapted to the colonization of open land, thereby reducing rotation time between successive harvests (Clapp, 2001). Currently, plantation forestry provides about 35% of wood consumption worldwide (Sedjo and Botkin, 1997), and although the approach has potential, native forests appear to suffer from even more vulnerable conditions as a result of the timelag between successive plantation harvests. They are now subjected to alternating periods of overexploitation and abandonment (Clap, 2001). Therefore, there is still a need to optimize the productivity of tree plantations to reduce these adverse effects.

To accomplish this, improvement of forest trees must be achieved through domestication, as in the case of agricultural crops (Campbell *et al*, 2003). However, traditional methods using breeding and selection of superior individuals are less amenable to tree improvement mainly because of the long maturation time of forest trees and their long life span and high heterozygosity (deleterious recessive alleles are conserved and result in inbreeding depression when breeding only a few individuals). With the advent of methods for *in-vitro* culture of trees from somatic tissue (clonal propagation via either organogenesis or embryogenesis) and molecular genetics, a new horizon for tree improvement came

¹ Canadian Forest Service. 2002-2003. The State of Canada's Forests, http://www.nrcan.gc.ca/cfs-scf/national/what-quoi/sof/sof02/pdf/UpFront_e.pdf (visited August 2004)

into being. It is now possible to detect superior genotypes for traits of high economic value through marker-assisted selection (quantitative genetics) and propagate them using *in-vitro* micro-propagation. Another possibility is to target specific traits by modifying the genetic background of trees by altering endogenous genes or incorporating new genes from related or unrelated species (genetic engineering), and then propagating the genetically modified organism (GMO) by *in-vitro* micro-propagation (Frankenhuyzen and Beardmore, 2004; Giri, Shyamkumar and Anjaneyulu, 2004). The first successful attempt to genetically engineer a forest tree was achieved with an angiosperm in 1987 by Fillatti *et al*, who had inserted an herbicide-tolerance gene in hybrid poplar using *Agrobacterium*-mediated DNA transfer. Six years later, the feat was accomplished for the first time with gymnosperms that include important timber species. Since gymnosperms are not easily infected by *Agrobacterium*, Ellis *et al* (1993), a team of Canadian researchers, used a different method that consisted of bombarding DNA-coated micro-projectiles to modify white spruce.

Key targeted traits in tree improvement include herbicide tolerance, insect and disease resistance, abiotic stress tolerance, wood quality (fibre content), high growth rate and short generation time. Despite important progress in achieving such high-value economic traits in transgenic trees, many scientific and social issues, most of which are related to safety, still impede their release for commercial applications. The problem lies in the lack of a clear understanding of the risks posed by the introduction of transgenes in nature, especially in trees due to their long lifespan. Safety issues include the spread of transgenes into wild relatives (vertical gene flow) without known potential adverse effects once in the genetic background of the natural population (variability in the effect of the transgene in different genotypes); spread of transgenes into unrelated organisms (horizontal gene flow), such as bacteria who could develop resistance to our fighting methods (resistance to herbicides); and selective advantage of transgenic trees over wild individuals, resulting in the replacement of the natural population with subsequent inbreeding depression or lack of adaptive potential due to the low genetic diversity of transgenic trees (clones) (Frankenhuyzen and Beardmore, 2004).

These considerable risks resulted in public fear of transgenic trees, and scientists now need to seek alternatives to provide high-yield forestry that is economically, ecologically and socially viable. One targeted solution consists of developing sterile transgenic trees that could be released in plantations without the risk of escapes. Although this approach is possible, the problem of reversibility of sterility still needs to be solved. Other alternatives include developing new breeding schemes to improve the quality and growth of trees, fighting pests and diseases using their natural enemies (biological control) such as bacteria (*Bacillus thuringiensis*) or insect viruses (Ebling, 2004) and exploiting lignin-degrading fungus rather than chemical products or transgenic trees to improve the pulping process (Kang *et al*, 2003). However, more knowledge of the genetics of forest trees and other interacting organisms is required to achieve these solutions, which require additional basic research in forest genomics.

The policy framework of the Canadian Biotechnology Strategy, renewed in 1998, recognizes that strong efforts in genomics research are necessary for Canada to become a world leader in this important emerging field. Forestry in particular has been identified as one of the key sectors, and the

international scientific community increasingly acknowledges that biotechnology can be fundamental to the development of tailored solutions to specific forestry issues. Taking this important information into consideration, the federal government has, since 1999, invested \$115 million in intramural research, out of which forest genomics initiatives received \$11 million through Natural Resources Canada's Canadian Forest Service. This funding allowed the Canadian Forest Service to increase resources and develop a strong forest genomics program. Other federal investments support genomics research in universities through the Natural Science and Engineering Research Council, the Canadian Foundation for Innovation, and the creation of Genome Canada which to date has a funding envelope of \$435 million. In addition, within the scope of its programs, Genome Canada has invested considerable amounts of money to bring Canada to the forefront of international research in an increasingly important sector of its economy. Until now, funding from Genome Canada in forest genomics has amounted to \$14 million, supporting four large scale projects across the country. Including other sources of funding, the total amount of money devoted to forest genomics in Canada totals \$28 million². In 2003, Genome Canada's efforts led to the establishment of an international collaboration program with Sweden with the goal of promoting exchange and greater achievement. Expected outcomes for both countries are "the development of solutions to protecting and maximizing the value of our forests, leading to new sustainable strategies for breeding, nurturing and exploiting trees"³.

In order to determine where Canadian researchers currently stand in respect to the rest of the international community and as a strategic tool to evaluate their scientific program, the Canadian Forest Service has mandated Science-Metrix to perform an analysis of forest genomics. Its aim is to study the evolution of the field at both the Canadian and international levels and to benchmark Canada's output against that of other leading countries in the field. In order to present a snapshot of scientific activity in the field of forest genomics worldwide, an analysis of publication data was conducted, providing an overview of the global rate of growth of scientific papers written in the field of forest genomics at the world level (Section 2) and a benchmarking of leading countries according to their scientific output (Section 3).

² Canadian Forest Service. 2002-2003. The State of Canada's Forests, http://www.nrcan.gc.ca/cfs-scf/national/what-quoi/sof/sof02/pdf/UpFront_e.pdf (visited August 2004)

³ Genome Canada, <http://www.genomecanada.ca/Gcmedia/communiquesPresse/indexDetails.asp?id=253&l=e> (visited August 2004)

2 Global trends in forest genomics

The number of scientific papers published in genomics as a whole increased rapidly at the world level from about 39,000 papers in 1992 to about 55,000 papers in 1998, at which point output levelled off. It was only in 2003 that worldwide output increased again, reaching a production of almost 60,000 papers per year for an overall increase of nearly 60% over the 12-year period (Figure 1). The significant growth in scientific production in genomics has translated into a considerable increase, although less substantial, of genomics' share of total scientific production. The percentage of genomics papers in the SCI Expanded database increased from 6.2% in 1992 to 7.6% in 1998, about 25% growth (see right y-axis, Figure 1). However, the slowdown in the increase of scientific production following 1998 resulted in a slight decrease of genomics' share of total scientific production, which decreased to 7.3% in 2003.

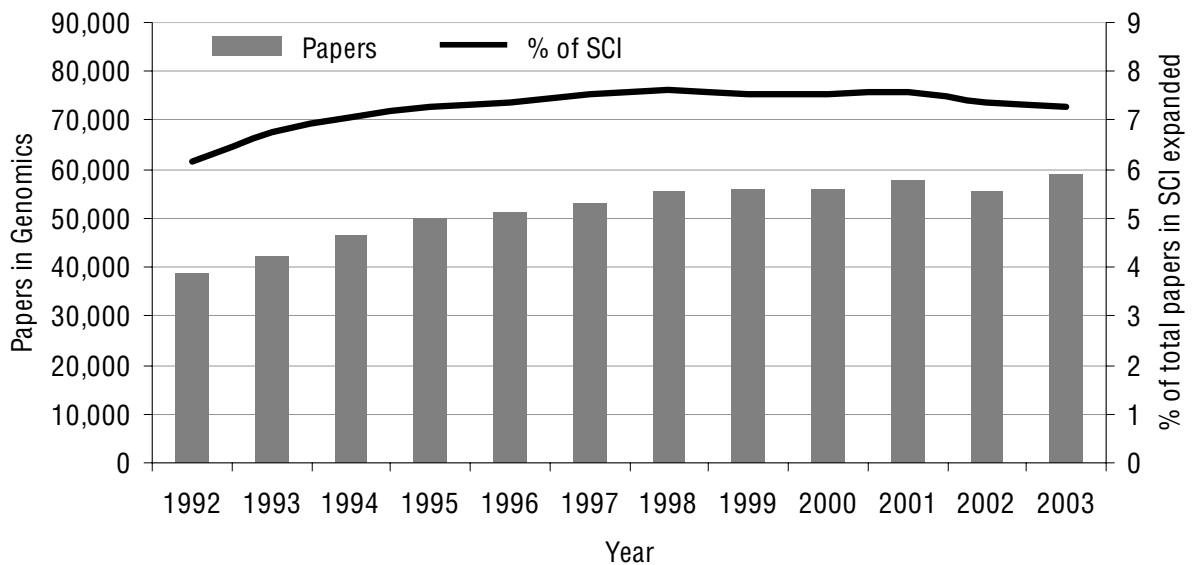


Figure 1 Papers in genomics in SCI Expanded, 1992–2003

Source: Compiled by Science-Metrix from SCI Expanded (Thomson ISI).

Within genomics, forest genomics is a small niche with a scientific production accounting for about 1% of the worldwide output in genomics overall in 2003 (see right y-axis, Figure 2). Although more uneven than in genomics, the growth of scientific output in forest genomics over the observed time-frame has also been substantial, with the number of papers per year increasing from 257 in 1992 to 542 in 2003, an increase of 110%.

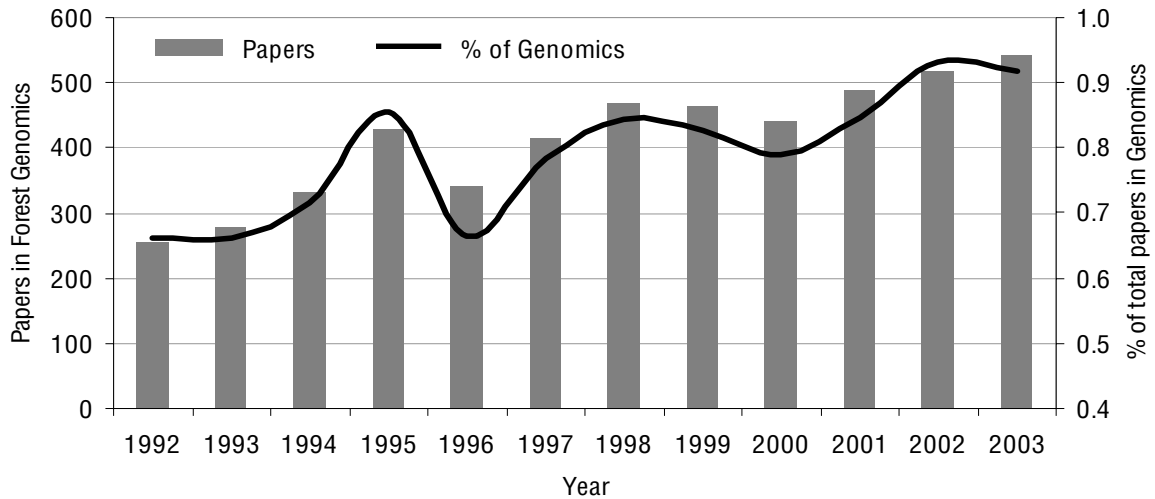


Figure 2 Papers in forest genomics in SCI Expanded, 1992–2003

Source: Compiled by Science-Metrix from SCI Expanded (Thomson ISI).

Figure 3 presents the top 30 countries in terms of scientific output in forest genomics. It is noteworthy that the Netherlands, Switzerland, Belgium and the Russian Federation, which rank 8th, 12th, 13th and 14th in terms of the number of papers in genomics overall (Archambault, Côté and Bertrand, 2003), do not fall among the 15 leaders in forest genomics. Conversely, Finland, the Republic of Korea, India and Brazil, which rank 17th, 19th, 20th and 23rd in number of papers in genomics overall (Bertrand and Côté, 2004) belong to the leading countries in forest genomics. The 15 leaders in scientific output in forest genomics will be retained for the following benchmarking analysis.

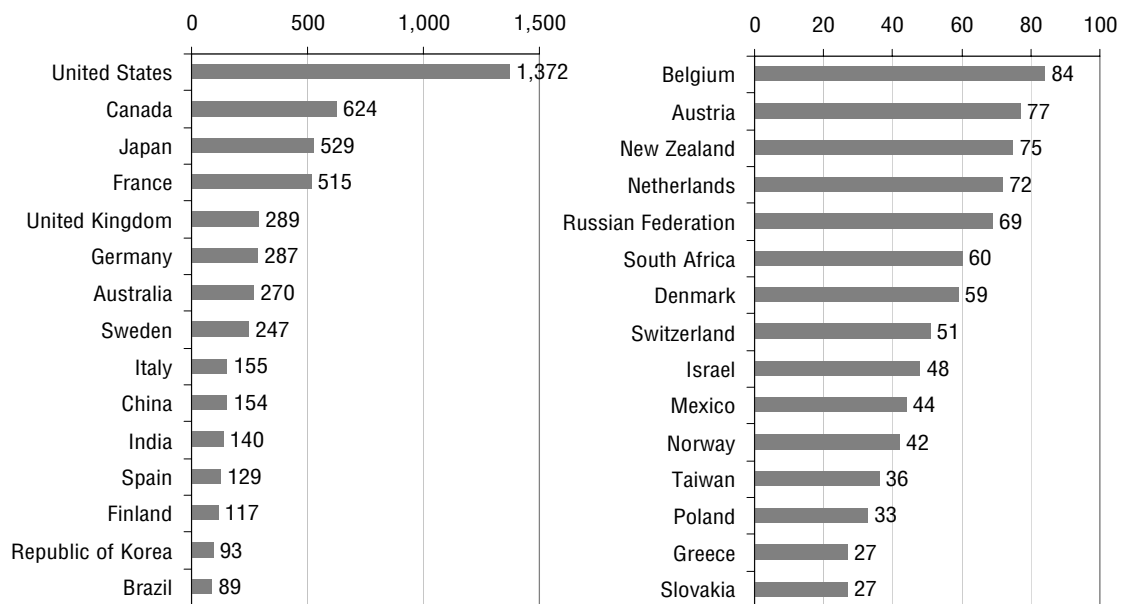


Figure 3 Top 30 countries by number of papers in forest genomics, 1992–2003

Source: Compiled by Science-Metrix from SCI Expanded (Thomson ISI).

3 Benchmarking leading countries in forest genomics

In order to benchmark the scientific output of the 15 leaders in forest genomics, comparable statistics across countries were used:

- Number of papers;
- Number of papers per capita;
- Specialization index (SI);
- Average of relative citations (ARC).

The indicators were subsequently combined in a multicriteria benchmarking of the 15 leaders. The present analysis is based on a 12-year period, which, in turn, is divided into four three-year periods to examine the field's evolution in the leading countries. Please refer to the Appendix for methods used in this scientometric analysis.

3.1 Number of papers in forest genomics

The US dominates forest genomics in absolute output, holding about 30% of the world's scientific production with 1,372 papers published over the 12-year period studied (Table I), and rank 1st in each of the time periods analysed (Table II). With 624 papers, about half as much as the US' output, Canada ranks 2nd overall in the period spanning 1992 to 2003 and in each of the four three-year periods. Japan and France follow closely with respectively 529 and 515 papers. Japan and France exchanged ranks over the four three-year periods, with France finishing in front of Japan in the 2001–2003 period. The UK, Germany, Australia and Sweden follow with similar output over the whole period, having published respectively 289, 287, 270 and 247 papers.

Australia and Sweden overtook the UK and Germany in the last three-year period, finishing on a par in 5th place, while the UK and Germany both took 7th place. Following at some distance behind the top producers with less than 200 papers over the 12-year period are Italy, China, India, Spain, Finland, the Republic of Korea and Brazil. Among them, Brazil and India significantly increased their scientific output during the 2001–2003 period, both with an average annual growth of 24%. Sweden, Australia and the UK also experienced significant growth in the last three-year period, with average annual growth above 20%. Canada, which is in a very good position in terms of absolute scientific output, ranked 11th for average annual growth between 2001 and 2003.

Table I Number of papers in forest genomics by leading countries per three-year period, 1992–2003

Country	1992-1994	1995-1997	1998-2000	2001-2003	1992-2003
United States	301	347	368	356	1,372
Canada	132	161	152	179	624
Japan	89	117	151	172	529
France	64	134	142	175	515
United Kingdom	44	66	88	91	289
Germany	34	70	92	91	287
Australia	32	57	81	100	270
Sweden	46	50	51	100	247
Italy	14	26	54	61	155
China	9	13	51	81	154
India	16	32	32	60	140
Spain	3	19	39	68	129
Finland	15	23	38	41	117
Republic of Korea	3	13	43	34	93
Brazil	8	12	27	42	89
World	869	1185	1375	1548	4977

Source: Compiled by Science-Metrix from SCI Expanded (Thomson ISI).

Table II Ranking of leading countries according to their number of papers in forest genomics, 1992–2003

Country	1992-1994	1995-1997	1998-2000	2001-2003	1992-2003
United States	1	1	1	1	1
Canada	2	2	2	2	2
Japan	3	4	3	4	3
France	4	3	4	3	4
United Kingdom	6	6	6	7	5
Germany	7	5	5	7	6
Australia	8	7	7	5	7
Sweden	5	8	9	5	8
Italy	11	10	8	11	9
China	12	13	9	9	10
India	9	9	14	12	11
Spain	14	12	12	10	12
Finland	10	11	13	14	13
Republic of Korea	14	13	11	15	14
Brazil	13	15	15	13	15

Source: Compiled by Science-Metrix from SCI Expanded (Thomson ISI).

3.2 Number of papers in forest genomics per capita

When considering each country's population size, larger countries with high absolute production such as the US, Japan, France, the UK and Germany are often overtaken by smaller countries with lower scientific production. Sweden and Finland, who rank 8th and 13th in absolute output, rank 1st and 2nd in output per capita, with 23 and 19 papers per ten million inhabitants per year (Table III and Table IV). Canada who was 2nd in production per capita in the 1992–1994 and 1995–1997 periods was surpassed by Finland in the last periods. Thus, Canada finished in 3rd place with an overall rank of 3rd for the 12-year period and an average of 17 papers per ten million inhabitants per year. The only other country to publish more than one paper per million inhabitants over the whole period was Australia in 4th place, with 1.2 papers per million inhabitants per year.

Among the remaining countries, France ranks 5th, the US 6th, the UK 7th, Japan 8th, Germany 9th, Spain 10th, Italy 11th, the Republic of Korea 12th, Brazil 13th, India 14th and China 15th. Countries' ranks over time were fairly stable, with the exception of that of the US, which dropped from 5th place in the 1992–1994 period to 9th place in the 2001–2003 period, and Spain, which started in 11th place in the first period and finished 6th in the last period (Table IV).

Table III Number of papers in forest genomics by leading countries per ten million inhabitants per year (annual average of periods), 1992–2003

Country	1992-1994	1995-1997	1998-2000	2001-2003	1992-2003
Sweden	17.58	18.84	19.17	37.55	23.28
Finland	9.86	14.96	24.56	26.36	18.94
Canada	15.21	17.93	16.37	18.70	17.05
Australia	6.02	10.36	14.24	17.03	11.91
France	3.70	7.65	8.01	9.73	7.27
United States	3.85	4.29	4.39	4.12	4.17
United Kingdom	2.53	3.75	4.95	5.06	4.07
Japan	2.38	3.10	3.98	4.51	3.49
Germany	1.40	2.85	3.74	3.68	2.92
Spain	0.25	1.59	3.25	5.64	2.69
Italy	0.82	1.51	3.12	3.51	2.24
Republic of Korea	0.23	0.95	3.06	2.36	1.65
Brazil	0.17	0.24	0.52	0.78	0.43
India	0.06	0.11	0.11	0.19	0.12
China	0.03	0.04	0.14	0.21	0.10
World	0.52	0.69	0.76	0.83	0.70

Source: Compiled by Science-Metrix from SCI Expanded (Thomson ISI). Population data were obtained from the U.S. Census Bureau.

Table IV Ranking of countries according to number of papers in forest genomics per million inhabitants (annual average of periods), 1992–2003

Country	1992-1994	1995-1997	1998-2000	2001-2003	1992-2003
Sweden	1	1	2	1	1
Finland	3	3	1	2	2
Canada	2	2	3	3	3
Australia	4	4	4	4	4
France	6	5	5	5	5
United States	5	6	7	9	6
United Kingdom	7	7	6	7	7
Japan	8	8	8	8	8
Germany	9	9	9	10	9
Spain	11	10	10	6	10
Italy	10	11	11	11	11
Republic of Korea	12	12	12	12	12
Brazil	13	13	13	13	13
India	14	14	15	15	14
China	15	15	14	14	15

Source: Compiled by Science-Metrix from SCI Expanded (Thomson ISI). Population data were obtained from the U.S. Census Bureau.

3.3 Specialization index - SI

The SI is an indicator of the intensity of research in a specific research area in a country relative to the intensity of research in the same research area in the world; a score above one means that a country specializes in the research area relative to the world. Over the 12-year period, Canada, Finland and Sweden were the most specialized countries in forest genomics, being more than twice as specialized as the world average in forest genomics (Table V) and occupying, respectively, 1st, 2nd, and 3rd place (Table VI). Australia, France, Brazil, the Republic of Korea, India, Japan and China also specialize in forest genomics albeit not as much as the three leading countries. The US, Germany and the UK are not specialized in this field, all producing a proportion of paper in forest genomics that is at least 30% less than the world figure (Table V). Among the 15 leading countries, Canada, the US and Japan have seen their SI decrease over time, while the SI of the remaining countries all increased. As a result, Canada lost some ground to Sweden and Finland, finishing in 3rd place in the 2001–2003 period (Table VI). The US position worsened more than that of Canada. Starting in 8th position in the 1992–2003 period, the US finished in 14th place in the last period, for an overall rank of 13. As for the number of papers per capita, Spain has made much headway over the 12-year period by increasing its SI from 0.16 in the 1992–1994 period to 1.43 in the 2001–2003 period. Indeed, it started in 15th place in the 1992–1994 period, and finished in 8th place in the 2001–2003 period, for an overall rank of 11th. In the last period, Sweden achieved the highest SI of all other countries in the 12-year period, being more than three times as specialized as the world average.

Table V Leading countries' specialization index in forest genomics per three-year period, 1992–2003

Country	1992-1994	1995-1997	1998-2000	2001-2003	1992-2003
Canada	2.99	2.80	2.43	2.59	2.66
Finland	2.18	2.21	2.89	2.72	2.60
Sweden	2.93	2.19	1.85	3.24	2.56
Australia	1.51	1.80	2.07	2.25	2.00
France	1.25	1.82	1.60	1.82	1.67
Brazil	1.26	1.16	1.63	1.82	1.63
Republic of Korea	0.71	1.17	2.06	1.04	1.43
India	0.79	1.28	1.09	1.66	1.27
Japan	1.22	1.13	1.18	1.19	1.18
China	0.64	0.49	1.20	1.17	1.06
Spain	0.16	0.65	1.00	1.43	0.98
Italy	0.47	0.58	0.97	0.92	0.80
United States	0.90	0.80	0.77	0.67	0.77
Germany	0.52	0.73	0.76	0.68	0.69
United Kingdom	0.57	0.61	0.67	0.65	0.63
World	1.00	1.00	1.00	1.00	1.00

Source: Compiled by Science-Metrix from SCI Expanded (Thomson ISI).

Table VI Ranking of leading countries according to their specialization index in forest genomics, 1992–2003

Country	1992-1994	1995-1997	1998-2000	2001-2003	1992-2003
Canada	1	1	2	3	1
Finland	3	2	1	2	2
Sweden	2	3	5	1	3
Australia	4	5	3	4	4
France	6	4	7	5	5
Brazil	5	8	6	6	6
Republic of Korea	10	7	4	11	7
India	9	6	10	7	8
Japan	7	9	9	9	9
China	11	15	8	10	10
Spain	15	12	11	8	11
Italy	14	14	12	12	12
United States	8	10	13	14	13
Germany	13	11	14	13	14
United Kingdom	12	13	15	15	15

Source: Compiled by Science-Metrix from SCI Expanded (Thomson ISI).

3.4 Average of relative citations - ARC

The ARC is an indicator of the number of times papers published by a country are cited relative to the world average in the field of genomics. A score above one indicates that a country's forest genomics papers are cited more often than papers in genomics on average at the world level. For the whole period studied, France, Spain, the US, the UK, Italy, Australia, Finland and Germany received more citations than the world average in forest genomics, while Japan Canada, China, the Republic of Korea, Brazil and India have ARCs below the world average (Table VII).

Some countries have made important progress over the 12-year period. In particular, Spain, which already made good progress with respect to the previous indicators, started in 14th place and gained 12 positions to finish 2nd behind France (Table VIII). France also improved its ranking notably, starting in 5th place and finishing in 1st. Germany gained 6 positions over the 12-year period, finishing in 3rd place (8th overall), and Italy gained 3 positions, finishing in 4th place (5th overall). Other countries, such as the US, the UK, Australia and Canada, lost some ground. Australia, which was 1st in the 1992–2003 period, finished in 5th place, for an overall rank of 6th. Canada whose forest genomics papers are less cited than the world average started in 6th place and dropped to 11th position in the 2001–2003 period, for an overall rank of 11.

Table VII Leading countries' average of relative citations in forest genomics per three-year period, 1992–2003

Country	1992-1994	1995-1997	1998-2000	2001-2003	1992-2003
France	1.04	1.26	1.31	1.52	1.33
Spain	0.28	0.87	1.33	1.47	1.31
United States	1.24	1.32	1.28	1.21	1.27
United Kingdom	1.22	1.29	1.20	1.30	1.26
Italy	0.99	0.83	1.29	1.43	1.24
Australia	1.40	1.08	1.13	1.31	1.22
Finland	0.38	1.47	1.34	1.03	1.14
Germany	0.69	0.90	1.00	1.47	1.09
Sweden	0.83	1.00	1.04	1.09	1.00
Japan	1.11	0.82	1.08	0.98	0.99
Canada	1.04	0.96	0.93	0.95	0.96
China	0.65	0.08	0.99	0.92	0.85
Republic of Korea	0.00	0.63	0.68	0.80	0.70
Brazil	0.58	0.51	0.72	0.36	0.51
India	0.41	0.32	0.35	0.67	0.48
World	1.00	1.00	1.00	1.00	1.00

Source: Compiled by Science-Metrix from SCI Expanded (Thomson ISI).

Table VIII Ranking of leading countries according to their average of relative citations in forest genomics, 1992–2003

Country	1992-1994	1995-1997	1998-2000	2001-2003	1992-2003
France	5	4	3	1	1
Spain	14	9	2	2	2
United States	2	2	5	7	3
United Kingdom	3	3	6	6	4
Italy	7	10	4	4	5
Australia	1	5	7	5	6
Finland	13	1	1	9	7
Germany	9	8	10	3	8
Sweden	8	6	9	8	9
Japan	4	11	8	10	10
Canada	6	7	12	11	11
China	10	15	11	12	12
Republic of Korea	15	12	14	13	13
Brazil	11	13	13	15	14
India	12	14	15	14	15

Source: Compiled by Science-Matrix from SCI Expanded (Thomson ISI).

3.5 Multicriteria analysis

The multicriteria ranking provides an overall assessment of the 15 leading countries in forest genomics. The criteria considered include the number of papers, papers per capita, specialization index and average of relative citations.

There were no clear leaders when considering all three-year periods (Table IX). However, France and Canada undoubtedly played a major role in the development of the field, ranking 1st and 2nd overall in the 12-year period. Canada was 1st in both the 1992–1994 and 1995–1997 periods. However, Canada subsequently slipped to 2nd place in the 1998–2000 period, on a par with France, and to 4th in the last period, behind France (1st), Sweden (2nd) and Australia (3rd). The diminishing performance of Canada is due in part to a specialization that gets lower over time but mainly to a relatively paltry performance, which is worsening over time, in terms of citations received by Canadian papers relative to the performance of other leading countries.

Overall, Sweden and Australia shared 3rd place. The US, which ranked 5th overall, lost ground over time, starting 2nd on a par with Sweden and finishing 7th on a par with Japan. Finland is 6th, Japan 7th, the UK 8th, Spain 9th, Germany and Italy both 10th, the Republic of Korea 12th, China 13th, and Brazil and India both 14th. The progress made by Spain has been tremendous. The country improved its position in forest genomics with respect to each indicator. According to the multicriteria ranking, Spain started last in the 1992–1994 period, then moved to 10th place, on a par with India in the 1995–1997 period, then to 9th place, on a par with Italy in the 1998–2000 period and finally to 6th place in the last period.

Table IX Multicriteria ranking of leading countries in forest genomics per three-year period, 1992–2003

Rank	1992-1994	1995-1997	1998-2000	2001-2003	1992-2003
1	Canada	Canada	Finland	France	France
2	Sweden (2)	France	Canada (2)	Sweden	Canada
3	United States (2)	Finland	France (2)	Australia	Australia (3)
4	Australia	Sweden	Australia	Canada	Sweden (3)
5	France	United States	Sweden	Spain	United States
6	Japan	Australia	United States	Finland	Finland
7	United Kingdom	United Kingdom	Japan	Japan (7)	Japan
8	Finland	Japan	United Kingdom	United States (7)	United Kingdom
9	Germany	Germany	Italy (9)	Germany	Spain
10	Brazil (10)	India (10)	Spain (9)	United Kingdom	Germany (10)
11	Italy (10)	Spain (10)	Germany	Italy	Italy (10)
12	India	Republic of Korea	Republic of Korea	China	Republic of Korea
13	China	Italy	China	Brazil	China
14	Republic of Korea	Brazil	Brazil	India	Brazil (14)
15	Spain	China	India	Republic of Korea	India (14)

Source: Compiled by Science–Metrix from SCI Expanded (Thomson ISI).

4 Analysis of Strengths and Weaknesses

The forest genomics dataset was divided into ten specialties in order to pinpoint the areas of research in which leading countries have strengths and weaknesses. The specialties are:

- **Tree genomics**, involving population and quantitative genetics studies of trees and the genetic engineering of trees, is the main specialty, with 3,285 papers published at the world level between 1992 and 2003 (Table X). The key goal of studies in tree genomics is to enhance tree performance to obtain high-yield plantations and therefore ease pressure on native forests.
- The **insect genomics** specialty follows in terms of the number of papers published over the period studied, with 427 papers focusing on the genetics of tree pests.
- The **bacteria genomics** specialty, which takes 3rd place with 365 papers, involves studying the genetics of tree diseases caused by bacteria and the genetics of bacteria which attack tree pests with the goal of developing efficient biological control agents.
- The **fungus genomics** specialty, which takes 4th place with 323 papers, involves studying the genetics of tree diseases caused by fungi.
- The **virus genomics** specialty, which takes 5th place with 206 papers, involves studying the genetics of tree diseases caused by viruses and the genetics of viruses which attack tree pests with the goal of developing efficient biological control agents.
- Scientific papers in the **pulp & paper** specialty focus on the genetics of fungi capable of degrading lignin and leaving behind the cellulose used in the production of paper and on the genetic engineering of trees to facilitate the pulping process. Pulp & paper ranks 6th with 160 papers.
- The **symbiosis genomics** specialty ranks 7th with 95 papers published over the 12-year period and covers genetic studies of symbiotic fungi specific to trees.
- The **saprophytism genomics** specialty also ranks 7th, with 95 papers published over the 12-year period and comprises genetic studies of saprophytic fungi specific to trees.
- The **nematode genomics** specialty, which takes 9th place with 12 papers, involves studying the genetics of tree diseases caused by nematodes.
- The **plant genomics** specialty, which takes 10th place with 9 papers, involves studying the genetics of tree diseases caused by plants.

Over the four three-year periods, the tree genomics, insect genomics, fungus genomics and virus genomics specialties were the most stable, growing steadily without dropping. Moreover, these specialties experienced high annual average growth over the 12-year period (fungus genomics: 183%; virus genomics: 155%; tree genomics: 130%; insect genomics: 86%). Although the growth of symbiosis genomics stopped in the last three-year period, it experienced the highest annual average growth over the 12-year period (400%). The evolution in the number of papers in the remaining specialties was less coherent, with alternating periods of growth and decline. Among them, the pulp & paper and bacteria genomics specialties had positive annual average growth of 64% and 42%, respectively. The nematode genomics (-67%), plant genomics (-50%) and saprophytism genomics (-44%) specialties had negative annual average growth over the 12-year period.

Table X Number of world papers in forest genomics by specialty, 1992–2003

Specialty	1992-1994	1995-1997	1998-2000	2001-2003	1992-2003
Tree genomics	544	772	905	1,064	3,285
Insect genomics	70	107	135	115	427
Bacteria genomics	94	95	84	92	365
Fungus genomics	57	74	93	99	323
Virus genomics	34	51	52	69	206
Pulp & paper	34	35	46	45	160
Symbiosis genomics	6	17	36	36	95
Saprophytism genomics	19	33	21	22	95
Nematode genomics	8	0	2	2	12
Plant genomics	3	1	1	4	9

Source: Compiled by Science-Metrix from SCI Expanded (Thomson ISI).

Figure 4 presents the strengths and weaknesses of the leading countries considering the specialties within forest genomics. In this case, data for the whole 12-year period was considered in aggregate in order to obtain large enough numbers for the computation of the two indicators used in the analysis (SI and ARC). For a specialty to be considered, it was required that at least ten papers be published over the period studied. The threshold was set at a relatively low value to ensure that at least two specialties remained for each of the countries analyzed (the top 8 on the multicriteria analysis). Nevertheless, caution should be exercised while interpreting the results when fewer than 30 papers have been published in a specialty over the 12-year period.

The strongest specialties of these countries can be found in the top right quadrant (high level of specialization, high level of citation). Conversely, the weakest specialties are to be found in the lower left-hand side quadrant.

France, which ranks 1st in the multicriteria analysis, is strong in tree genomics (385 papers), in bacteria genomics (41 papers) and symbiosis genomics (29 papers). Although France's SI in pulp & paper (15 papers) is slightly below the world average, the country ARC in this specialty is about 1.5 times the world average so that it can be considered a potential strength. France's main weaknesses lie in insect genomics (22 papers) and fungus genomics (16 papers).

Within Canada (2nd place overall), the main strength lies in fungus genomics (59 papers, 2nd rank). Although the impact of Canadian research in tree genomics, as inferred from ARC, is the lowest among the 8 leaders and slightly below the world average, the country is more specialized than the world average and ranks 2nd in terms of number of papers produced in this specialty (443 papers). In addition, tree genomics is Canada's number one specialty in terms of the absolute production in forest genomics and can thus be considered an important strength of Canada in the field. Potential strengths for Canada include the insect genomics and virus genomics specialties. In insect genomics (40 papers, 3rd rank), Canada's impact (ARC) is more than 1.5 times the world average (the highest among specialties for Canada), but its SI is below the world average. In virus genomics (47 papers, 2nd

rank), the picture is reversed with Canada's SI being 1.5 times higher than the world average and its ARC being less than 1.5 times the world average. Canada's main weakness lies in bacteria genomics (17 papers, 7th rank).

Australia and Sweden rank 3rd on a par in the multicriteria analysis and they are both strong in tree genomics (214 and 190 papers, respectively). Sweden is also strong in fungus genomics (27 papers). Australia has a potential strength in insect genomics (17 papers) where its impact is about 1.5 times the world average, but should increase its emphasis in this specialty to increase its SI. Sweden's potential strength lie in pulp & paper (16 papers) where its SI is about twice that of the world average and its ARC is only slightly below the world average. Australia has a weakness in fungus genomics (12 papers).

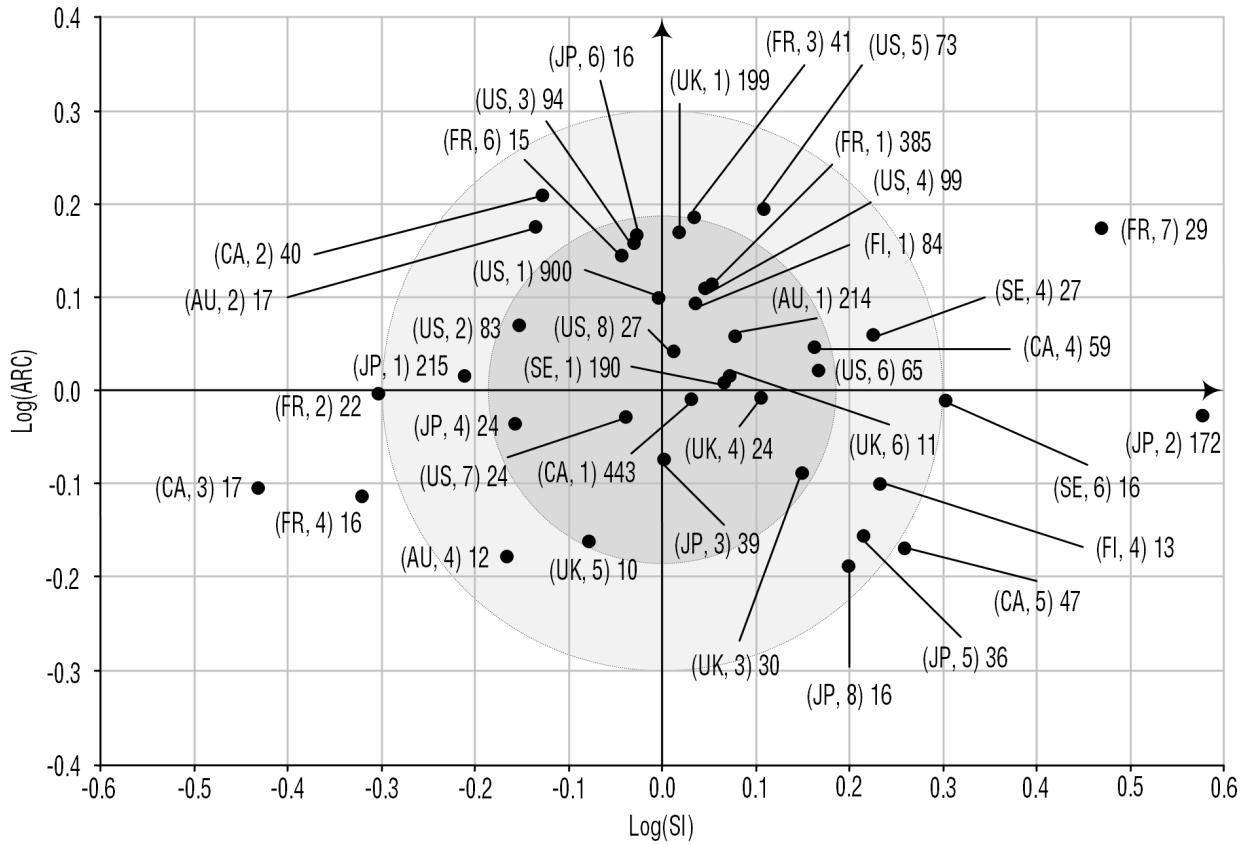
The US (5th place overall) performs very well in 5 out of 10 specialties. These are tree genomics (900 papers), fungus genomics (99 papers), virus genomics (73 papers), pulp & paper (65 papers) and saprophytism genomics (27 papers). It has a potential strength in bacteria genomics (94 papers) and in insect genomics (83 papers), and a weakness in symbiosis genomics (24 papers).

Finland (6th place overall) has a strength in tree genomics (84 papers) and a potential strength in fungus genomics (13 papers). In order to build on this potential strength, Finland would need to increase the impact of its research in fungus genomics.

Japan (7th overall) performs well in insect genomics (172 papers) and in pulp & paper (16 papers). It has potential strengths in virus genomics (36 papers) and saprophytism genomics (16 papers) in which it has high SIs, but relatively low ARCs. Its main weaknesses are in tree genomics (215 papers), bacteria genomics (39 papers) and fungus genomics (24 papers).

Finally, the UK come last among the top 8 leaders. The country is strong in tree genomics (199 papers), in pulp & paper (11 papers) and in fungus genomics (24 papers). It has a potential strength in bacteria genomics (30 papers) and a weakness in virus genomics (10 papers).

Benchmarking of Forest Genomics in Canada and Other Leading Countries



Legend			
Country code		Specialty Code	
AU: Australia	JP: Japan	1 : Tree genomics	6 : Pulp & paper
CA: Canada	SE: Sweden	2 : Insects	7 : Symbiosis
FI: Finland	UK: United Kingdom	3 : Bacteria	8 : Saprohytism
FR: France	US: United States	4 : Fungi	9 : Nematodes
		5 : Viruses	10: Plants

Note: Indicators are log transformed such that the origin (0, 0) represents the world average. For both indicators, a value above 0 means that the score is higher than the world average, while a value below 0 means that the score is inferior to the world average. Consequently, strengths are to be found in the upper right quadrant, while weaknesses are to be found in the lower left quadrant. The inner dotted circle crosses the x or y-axis at a value that corresponds to a SI or ARC that is 1.5 times the world average. The outer dotted circles cross the x and y-axis at a value corresponding to a SI or ARC that is twice that of the world average. Strengths and weaknesses of countries are represented by a dot with a tag. The tag (CA, 1) specifies the country name and the specialty. The number following the tag indicates the absolute production of scientific papers by the country in the corresponding specialty.

Figure 4 Leading countries' strengths and weaknesses in forest genomics, 1992–2003.

Cartesian plan combining specialization index (SI) and average of relative citations (ARC) by specialty.

Source: Compiled by Science-Metrix from SCI Expanded (Thomson ISI).

5 Conclusion

Forests lie at the heart of Canada's economy, and their exploitation involves competing timber- and non-timber based activities. Indeed, Canada's forests form the basis of an industry worth \$74 billion annually that provides direct employment to about 350,000 Canadians and is a stepping stone of the Canadian tourism industry, which is also worth several billion dollars a year⁴. In addition, forest conservation has become a priority on the national agenda with increased public awareness of the problems resulting from deforestation. In order to cope with these competing activities, the federal government is currently working at improving its strategy for sustainable forest management.

Although forest genomics is a small niche within the field of genomics (about 1% of the worldwide output in genomics in 2003), its growth has been substantial over the observed time-frame, with the number of papers per year more than doubling from 257 in 1992 to 542 in 2003. As genomics appears to be one of the most promising avenues to develop tailored solutions to specific forestry issues, the federal government has invested \$28 million since 1999 in forest genomics through Natural Resources Canada's Canadian Forest Service and other R&D vehicles.

Canada has traditionally been a major player in the development of innovative genomics approaches to improve the management of forests. Indeed, using a multicriteria analysis, this study has revealed that Canada ranked 2nd behind France for its scientific output in forest genomics for the period spanning 1992 to 2003. However, Canada lost some ground over the period, starting 1st in the 1992-1994 period and finishing 4th in the 2001-2003 period. When dissecting the field of forest genomics, Canada excels in tree genomics and genomic studies of parasitic fungi of trees, whereas it lags behind other countries in genomic studies of parasitic bacteria of trees or tree pests. Canada also has potential strengths in genomic studies of tree pests (insects) and viruses of trees or tree pests. Overall, Canada's strengths are its considerable scientific output (2nd in number of papers), large number of papers per capita (3rd) and strong specialization in this area (1st in terms of specialization index). Its main weakness relates to the relatively small number of citations received by its scientific papers. Hence, in order to maintain its leadership in an expanding field that promises to solve problems in a crucial sector of our economy, Canada should at the very least maintain its current level of funding, which is substantially higher than at the time when Canada embarked on its downward trend in terms of scientific impact. This will ensure that Canadian researchers continue to have access to the modern equipment that is essential when pursuing research in a field that is undergoing rapid technological development.

Given the tremendous amount of work required in genome sequencing and functional analysis, collaboration between countries with similar research interests is a sure way to cut development time and to lower costs. Thus, it would be of interest to map international collaboration and to identify potential collaborators with whom mutually beneficial alliances could be developed. In addition, a more in-depth study conducted at various levels of aggregation (e.g. provinces, institutions, researchers) could explain the dynamics of research in forest genomics within Canada and help identify potential centres of excellence.

⁴ Canadian Forest Service. 2002-2003. The State of Canada's Forests, http://www.nrcan.gc.ca/cfs-scf/national/what-quoi/sof/sof02/pdf/UpFront_e.pdf (visited August 2004)

References

- Archambault É., Côté G. and Bertrand F. 2003. Benchmarking of Canadian Genomics, 1991–2002. Prepared for Genome Canada by Science-Metrix, 49 pages.
- Bertrand F. and Côté G. 2004. Benchmarking genomics and health biotechnology in seven developing countries, 1991-2002 - Brazil, Cuba, China, Egypt, India, Republic of Korea and South Africa. Prepared for the University of Toronto Joint Center for Bioethics (JCB) by Science-Metrix, 78 pages.
- Clapp R.A. 2001. Tree Farming and Forest Conservation in Chile: Do Replacement Forests Leave Any Originals Behind? *Society and Natural Resources* 14: 341–356.
- Campbell M.M., Brunner A.M., Jones H.M. and Strauss S.H. 2003. Forestry's Fertile Crescent: The Application of Biotechnology to Forest Trees. *Plant Biotechnology Journal*, 1: 141–154.
- Ebling P.M. 2004. Comparative Activity of Choristoneura Fumiferana Nucleopolyhedrovirus Propagated in Different Hosts. *Pest Management Science*, 60; 631–638.
- Ellis D.D., McCabe D.E., McInnis S., Ramachandran R., Russel D.R., Wallace K.M. and Martinell B.J. 1993. Stable Transformation of *Picea glauca* by Particle Acceleration. *Biotechnology* 11: 84–89.
- Fillatti J.J., Sellmer J., McCown B., Haissig B., Comai L. 1987. *Agrobacterium*-mediated Transformation and Regeneration of *Populus*. *Mol. Gen. Genet.*, 206: 192–199.
- Frankenhuysen K.V. and Beardmore T. 2004. Current status and environmental impact of transgenic forest trees. *Canadian Journal of Forest Research*, 34: 1163–1180.
- Giri C.C., Shyamkumar B., Anjaneyulu C. 2004. Progress in Tissue Culture, Genetic Transformation and Applications of Biotechnology to Trees: An Overview. *Trees*, 18: 115–135.
- Kang K.Y., Jo B.M., Oh J.S. and Mansfield S.D. 2003. The Effects of Biopulping on Chemical and Energy Consumption during Kraft Pulping of Hybrid Poplar. *Wood and Fiber Science*, 35: 594–600.
- Kremen C., Niles J.O., Dalton M.G., Daily G.C., Ehrlich P.R., Fay J.P., Grewal D. and Guillery R.P. 2000. Economic Incentives for Rain Forest Conservation across Scales. *Science*, 288: 1828–1832.
- Marcoux A. 2000. Population and the environment: a review and concepts for population programmes. Part III: Population and deforestation. *Food and Agricultural Organization of the United Nations*, <http://www.fao.org/sd/wpdirect/WPan0050.htm>.
- Sedjo R.A. and Botkin D. 1997. Using Forest Plantations to Spare Natural Forests. *Environment*, 39: 14–20.

Appendix: Methods Used in Scientometric Analyses

Dataset

This scientometric study is based on the use of the Thomson ISI Science Citation Index Expanded database (SCI Expanded). The statistics are drawn from four document types that are considered to be original contributions to scientific knowledge: articles, notes, reviews and conference proceedings. The tables presented in this report refer to these four document types as "papers".

A subset of papers on forest genomics has been selected in an already defined subset of papers on genomics. The genomics subset is defined as follow:

Genomics - Genomics was defined as the study of genomes, specific genes and their function. This definition includes classical genomics, which focuses on gene and DNA sequencing, and functional genomics, which focuses on gene interaction and evolution. It also includes research that uses these genome-based technologies to effectively pursue questions in the life sciences.

Operationally, the construction of the genomics dataset is essentially based on the use of keyword-in-title searches. This keyword set has been fine-tuned over the years and was originally defined by experts named by Genome Canada and analysts now working for Science-Metrix. These keywords have been tried and tested. The search expression used selects papers in core and peripheral genomics (such as papers in molecular biology that touch upon the genome).

Forest Genomics - The set of papers on forest genomics was built by searching forest-specific keywords in titles of genomics papers. Extensive lists of trees, pests and other relevant keywords were used. Particular care was taken to ensure unbiased coverage for each country.

Each paper from the forest genomics set was categorized in specialties, again using keyword-in-title searches.

Caveats – The use of keyword searches limited to titles tends to underestimate scientific output, as some papers do not have a title that contains keywords specific to their respective field. However, because keywords-in-title are more subject-specific than keywords-in-abstract or keywords-in-full text, this strategy reduces the occurrence of false positives. The keyword-in-title method is adequate since the goal of this study is not to produce a census of papers in the field but rather to provide an unbiased comparison of the production of countries active in forest genomics. Nonetheless, SCI Expanded does provide the most extensive coverage of high-quality scientific research, because it indexes approximately 5,900 of the world's leading scholarly science and technical journals in more than 150 disciplines⁵. These journals are considered to be the most important peer-reviewed journals in their respective fields. They reflect significant scientific achievements and are the most widely cited journals in the world (over 80% of the world's citations). In addition, SCI Expanded is the only database that indexes the reference made by and the citation received by papers. This information is

⁵ Data derived from information prepared by the Institute for Scientific Information, Inc. (ISI, Philadelphia, Pennsylvania, USA). Copyright Institute for Scientific Information. All rights reserved.

invaluable in the benchmarking of scientific research. The choice of SCI Expanded over other databases (e.g. Biological Abstract, Agricola) is also advantageous in so far as that it allows for the opportunity of analysing collaboration between countries, because only Thomson ISI's databases traditionally comprised the address of every author listed in scientific papers.

Indicators

The resulting datasets were used to produce detailed statistics based on the following indicators:

Number of papers - Number of scientific papers written by authors located in a given geographical, sectoral, organizational or individual entity (e.g.: countries, cities or institutions).

Specialization index - This is an indicator of the intensity of research of a given geographic or organizational entity (e.g.: country) in a given research area (domain, field) relative to the intensity of the reference entity (e.g.: world) in the same research area. The specialization index can be formulated as follows:

$$SI = \frac{(X_s / X_T)}{(N_s / N_T)}$$

where

X_s = Papers from entity X in a given research area (e.g.: Canada in forest genomics)

X_T = Papers from entity X in a reference set of papers (e.g.: Canada in the whole SCI Expanded database)

N_s = Papers from the reference entity N in a given research area (e.g.: world in forest genomics)

N_T = Papers from the reference entity N in a reference set of papers (e.g.: world in the whole SCI Expanded database).

An index above 1 means that a given entity is specialized relative to the reference entity. Conversely, a score below 1 means that the given entity is not specialized relative to the reference entity.

For the current study, the specialization index as computed for forest genomics uses the whole SCI Expanded database as a reference set of papers. The specialization index as computed for subfields uses the forest genomics set of papers as a reference.

Average of relative citations – This is an indicator of the scientific impact of papers produced by a given entity. The number of citations received for each paper was counted for the year in which they were made and for the two subsequent years. For papers published in 1992, for example, citations received in 1992, 1993, and 1994 were counted. The only exceptions are 2002, which comprises a citation window of two years (2002 and 2003), and 2003, which contains a citation window of one year, since citation data was not yet available for the subsequent years. For a paper in a given subfield (reference subfields are those defined by CHI Research Inc for the National Science Foundation and used in the Science and Engineering Indicators), the citation count was then divided by the average

count of all papers in this subfield (reference set of papers = genomics dataset) to obtain a relative citation count (RC). The average of relative citations (ARC) of a given entity are computed using the average RC of each paper belonging to it. When the ARC is above 1, an entity scores better than the world in genomics on average; when it is below 1, an entity publishes papers that are not cited as often as the world in genomics on average.

Analysis of strengths and weaknesses

In order to identify countries' strengths and weaknesses, an analysis of scientific output was conducted for each of the ten specialties identified in forest genomics. For each specialty and country, the SI and ARC were calculated. For the SI, research areas are the specialties, and the reference set of papers is the whole forest genomics dataset. For the ARC, the reference set of papers is the whole genomics dataset as is the case for the whole field of forest genomics. For a specialty to be incorporated into the analysis for a given country, the country had to have published at least ten scientific papers in that specialty over the 12-year period. Only the top eight leaders were retained in the analysis of strengths and weaknesses.