

# We are IntechOpen, the world's leading publisher of Open Access books Built by scientists, for scientists

6,500

Open access books available

176,000

International authors and editors

190M

Downloads

Our authors are among the

154

Countries delivered to

TOP 1%

most cited scientists

12.2%

Contributors from top 500 universities



WEB OF SCIENCE™

Selection of our books indexed in the Book Citation Index  
in Web of Science™ Core Collection (BKCI)

Interested in publishing with us?  
Contact [book.department@intechopen.com](mailto:book.department@intechopen.com)

Numbers displayed above are based on latest data collected.  
For more information visit [www.intechopen.com](http://www.intechopen.com)



## Chapter

# Are There Adequate Incentives for Research and Innovation in the Plant Breeding Supply Chain?

*Predrag Rajsic, Richard Gray, Alfons Weersink  
and Istvan Rajcan*

## Abstract

The breeding supply chain has expanded with genomic technology to include basic research scientists and applied genomicists along with traditional plant breeders and farmers. Genomicists have focused on identifying specific DNA sequences or quantitative trait loci (QTL) that can be used as molecular markers. However, the use of molecular marker-assisted selection (MAS) by breeders in their programs requires the identified QTL to be reliably correlated with agronomically desirable traits. Replication research is critical for reducing the risk associated with the adoption of new marker-based (or QTL-based) selection strategies, but the applied scientists doing genomics research often do not have the incentives to do replication and other research required to verify the reliability of markers. The misalignment of incentives in the breeding supply chain can curtail the development of the projected advances in food production by genomics research. Using a sample of 24 genomic journals, we found more highly ranked journals tend to favor new research on identifying new QTL over replication research on previously identified QTL. Given that breeders will tend to adopt only those markers perceived to be reliable, the implicit lack of incentives for basic and applied genomic scientists to undertake replication research can impede agricultural innovation.

**Keywords:** genomic technology, plant breeding, replication research, QTL discovery

## 1. Introduction

The challenge of meeting the food and energy needs of a growing, increasingly wealthy, and global population from a finite and increasingly compromised resource base is formidable [1]. The hope of meeting the challenge rests with continued scientific progress and innovation [2]. Technological developments through the 20th century, largely fueled by publicly funded research programs, allowed the amount of agricultural output to nearly quadruple while the weighted index of real prices for 18 associated products has fallen by 75% [3]. Given the limited availability of untilled farmland and increasing agri-environmental constraints, future expansion will have to occur from increases in the productivity of cultivated land [4]. However, the rate of

yield growth has generally declined since 2000 as governments have shifted funding to other priorities [5].

Scientific discoveries in the field of genomics create the potential for substantial future yield increases. The scope for innovation associated with genomics and related fields has some dubbing the 21st century as “the century of biology” [6]. In agriculture, genomic science is rapidly expanding the pool of knowledge that can accelerate the development of improved crops and animals. To realize this potential, it is important to consider the process of innovation and, in particular, how scientific knowledge gets translated into new products and processes.

Genomic research has resulted in the sequencing and mapping of the DNA for most large commercial crops and most of the important livestock species [7]. To date, genomicists, working with breeders, have been able to identify thousands of specific DNA sequences or quantitative trait loci (QTL) as molecular markers associated with many important (phenotypic) crop traits. As a result of these discoveries, many breeders are now using molecular marker-assisted selection<sup>1</sup> (MAS) to augment phenotypic selection in their breeding programs. Importantly, this knowledge is only useful to breeders when the QTL are reliable (i.e., they are highly correlated with the trait), the trait is commercially (agronomically) important, and MAS is cost-effective to deploy. The process of successful innovation or deployment of genomic knowledge requires the coordination and the activity of basic research scientists and applied scientists working with breeders and seed firms, the commercial development of sequencing platforms and equipment, and in the end, the farmer adoption of the new genetics in the form of new crop variety or animal genetics.

The breeding processes involving genomic research can be considered within a supply chain framework involving four relatively distinct groups: (1) basic research scientists who make discoveries useful to applied genomicists, (2) applied genomicists who develop molecular markers by identifying the associations between genes and the expression of a given trait (3) breeders who use MAS in their breeding program, and (4) farmers who incorporate the new varieties into their operation to increase production [8]. For this supply chain to operate effectively, each link must be strong and securely connected to the adjoining links in the chain. Each link must have the required resources and incentives to produce what is needed for uptake downstream.

This brings us to the issue addressed in this paper. While breeders need reliable markers (QTL), the applied scientists doing genomics research often do not have the incentives to do replication and other research required to verify the reliability of markers. The misalignment of incentives in the breeding supply chain has the potential to curtail the development of the projected advances in food production made possible through genomics research.

In this paper, we consider how the development and use of MAS in plant breeding can be influenced by the metrics used to reward basic and applied genomic scientists. We show that scientists will be incentivized to focus on new QTL discoveries at the expense of verifying the previously discovered QTL through replication research if the metric used to reward these scientists is the relative rank of the scientific journal in which their article appears. Journals are ranked by the number of times a scientific article is cited by later scientific articles, and it is more prestigious for a scientist to publish in higher-ranked journals. Any bias in the type of articles accepted by more highly ranked journals would

---

<sup>1</sup> MAS is a form of genomic selection where a relatively small number of genetic markers are used in the selection process.

affect the type of research carried out. We will show that more highly ranked journals publish more articles on MAS discovery rather than verification. Given that breeders will tend to adopt only those markers perceived to be reliable, the implicit lack of academic reward for verification studies can distort incentives for basic and applied genomic scientists to undertake replication research, which, in turn, can impede agricultural innovation. Having highlighted this issue, we make a more general argument that applied genomic researchers have a unique role in innovation systems and, therefore, require a modified set of incentives rather than just relying on journal ranking alone.

The replicability crisis identified by Ioannidis [9], later documented by Open Science Collaboration [10], Baker [11], and Fanelli [12] compounds the incentive compatibility problem. Ioannidis [9] finds that most biometric studies may be false or an expression of a prevailing bias. Open Science Collaboration [10] conducted replications of 100 experiments published in three high-ranking psychology journals in 2008 and found that while 97% of the original experiments had statistically significant results only 36% of the replications had statistically significant results. In a survey of scientists' views of reproducibility by Baker [11], over 70% of the 1500 researchers surveyed reported that they have tried and failed to reproduce another scientist's experiments and 52%, 38%, and 7% of respondents stated that there is a "significant reproducibility crisis," "slight crisis," or "no crisis," respectively. Fanelli [12] concludes that although science may not be facing a reproducibility crisis, reproducibility is an important challenge that needs to be addressed. All these issues create further disincentives for researchers attempting replication research and have subsequent implications for the design of research funding programs.

The private rewards for discovery research, including prestige from publication in a higher quality journal, should be higher given the higher costs and skills to undertake such research. However, the idea of incentive misalignment stems from the public aspect of the benefits of replication research along the supply chain. Replication research and new research can be viewed as complementary inputs in productivity growth stemming from plant breeding. Thus, the public marginal benefits of some new/discovery research may be wasted if not coupled with replication research. Our research does not prove that the mix of replication and new research is suboptimal, but it does give some evidence that it might be. Further research is needed to assess potential efficiency gains from alternative mixes of new research and replication research.

The remainder of the paper is organized as follows. We begin with a conceptual framework that is informed by a description of the research and development required for the implementation of MAS in plant breeding. We also describe factors that breeders weigh in their adoption decision. Given the literature on supply chain management (SCM) and incentive alignment, we review previous studies that have examined agricultural innovation systems and incentives. Based on this literature and the description of the research and development required for the implementation of MAS in plant breeding, we develop a potential rationale that suggests that there may be a lack of incentives for replication research in genomics. To assess this rationale, we examine the statistical relationship between marker discovery versus verification and journal ranking. Building on the results of this analysis, we show that output metrics based solely on journal ranking, create a strong incentive for genomic scientists to focus on discovery rather than verification, and without verification, the MAS is less likely to be adopted. We conclude the paper with a discussion of the implications, arguing that funding bodies and public administrators interested in innovation should consider a broader range of metrics regarding replication research, particularly given the value of applied research to downstream users.

## **2. Conceptual framework**

### **2.1 The plant breeding supply chain and genomic selection**

The use of marker-assisted selection (MAS) has become an important tool for the breeders of many crops. Recent advances in genomics have linked particular DNA molecular markers to the phenotype or function of the plant. Breeders are increasingly using these molecular markers to help, select, and screen the plants for their breeding program. For example, if a wheat breeder is interested in developing varieties resistant to a particular strain of leaf rust, the molecular marker for rust resistance can be used as a screen to identify lines with this resistance trait, saving time, and, potentially, cost in the breeding process. This usefulness of MAS has created a demand for applied genomics science where researchers work to find molecular DNA markers associated with important phenotypic traits in crops.

The use of MAS in breeding programs relies on the existence of markers of interest, their reliability, and the breeders' cost of implementation. For many crops, there is already a long list of molecular markers, often numbering in the thousands [13]. A breeder must judiciously decide which markers to use in their breeding program because the total number of markers that can effectively be used for variable selection is limited by the effect on plant populations. If the traits exist in a single gene of a heterozygous parent, only half of their offspring will inherit the trait. As such, every time a marker is used for selection, it will reduce the size of eligible lines in the first generation (F1) by 50%. If a breeder uses four markers to select the first generation, only one in 16 of the F1 generation would have all four markers. If 10 markers were used, 1 in 1024 would meet the screen, eliminating 99.9% of the lines. As such, breeders must very carefully consider which markers and how many markers they are going to use in their breeding program.

The cost of implementing MAS will also affect both the number of markers and the stage of implementation in the breeding program [13]. One of the strongest drivers of increased MAS adoption has been the decrease in the per-sample cost of analysis using technology. Early technologies had limited scope and were time-consuming and labor-intensive, resulting in large costs per sample. When MAS application was expensive, breeders would either avoid using them, use them for the parent lines only, or use them later in the breeding process at F4 or F5 generation, after they had considerably narrowed the potential number of lines by other means. As the sample cost has fallen, more breeders are able to use MAS in the selection of F1 generation.

Marker reliability is critically important for breeder adoption. Markers are typically discovered through a statistical association of a particular phenotype and the genetic marker. If the marker and a single gene responsible for the trait are closely located on the same chromosome, there will be a high correlation between the existence of the marker and the phenotype produced by that gene. If, however, the marker is not close to the gene responsible for the phenotypic trait, the gene can exist in the absence of the marker (i.e., a type 1 error) and the marker can be present without the phenotypic trait (i.e., a type 2 error). A marker is considered reliable when both type 1 and type 2 errors are close to or equal to zero.

The economic costs of both type 1 and type 2 errors are important. In the case of type 1 error, selecting a line that does not have the desired trait increases the cost of all

subsequent downstream development, until the error is found. For example, if a program carries 20% of lines that do not have the desired trait, the costs per viable line would increase proportionally. With type 2 errors, rejecting a line that in fact does have the desired trait reduces the number of lines that could be used for subsequent selection. A 15% chance of a type 2 error at the F1 stage would require the breeder to have a population 15% larger to yield the same number of viable lines after the MAS. Given the economic importance of marker reliability, it is easy to understand why breeders often identify marker reliability as a key factor in the adoption decision.

The reliability of markers can be improved by further genomic research that tests the relationship between MAS and desired genes in a larger population while looking for more reliable markers at each stage. The actual gene (which is the perfect marker) can be identified often through a process of elimination and more recently can be verified through gene editing techniques. Given both the theoretical and empirical evidence pointing to the importance of marker reliability, research that verifies and improves marker reliability is valuable for downstream breeders and innovation outcomes.

The development and use of MAS is an excellent example of how certain advances in science can be mobilized for breeding, resulting in improved crop varieties and improved agricultural productivity. If one conceptualizes this as an innovation system involving components of basic science, genomics, applied science, breeder, and commercial use, many individuals and institutions are involved.

## **2.2 Agricultural innovation systems and incentives**

Well-functioning supply chains require communication, coordination, and an alignment of incentives for participants along the chain. While markets can play an important role in the coordination along a supply chain, markets also often fail, requiring other mechanisms including contracting or other forms of vertical integration to provide effective low-cost coordination. Firms or industries that are successful in developing cost-effective supply chains gain a competitive advantage and can prosper, while those that fail to efficiently achieve the coordination required, lose market share and can be driven out of the market. Porter [14] argues that communication and knowledge flows are key to innovation and the competitive advantage of nations.

An important branch of the SCM literature examines innovation systems that allow firms, industries, and countries to create and mobilize knowledge to increase productivity [15]. In our experience, agricultural research differs from many industries because of the extensive involvement of the public sector throughout the breeding supply chain from basic and applied research to crop breeding and agronomy to farm extension programs. However, many of the public-based systems have evolved to include private research and commercialization firms within the supply chain. A wide variety of institutional arrangements involving public-private partnerships exist in crop innovation systems globally [16]. In the case of MAS, much of the basic science, marker discovery, and verification continues to be done primarily by the public sector, with private firms increasingly engaged in breeding and variety development.

There are many examples of highly successful crop innovation systems that have linked cutting-edge genomics with the widespread deployment of MAS and other applied tools. In the United States, the United States Department of Agriculture (USDA) and other public scientists have undertaken the basic research and genomics that have supported the development of very large and sophisticated private

molecular breeding programs in soybeans and corn) [17]. In Canada, Genome Canada has supported many large crop genomics projects that are integrated into public, producer, or privately supported breeding programs. In France, BreedWheat, a large public-private research consortium made up of 14 organizations, has been collaborating since 2010 to undertake genomics research with the goal of supporting the development of wheat varieties [18].

There are also many examples where effective supply chains for MAS have failed to develop [19]. For example, in our own soybean breeding and genetics lab at the University of Guelph, we have developed many markers through graduate student projects that were published but never used in MAS (Rajcan, personal communication). These failures can be roughly attributed to a lack of resources and/or incentives required to support the genomic research necessary to develop reliable molecular markers. Generally, these failures tend to be associated with either minor crops or crops primarily grown in countries with limited public resources [19]. Other failures in MAS adoption have occurred where the markers exist but breeders lack the knowledge and/or cost-effective tools to undertake and deploy MAS.

A lack of coordination can impair a supply chain even when adequate resources exist. One classic failure occurred in the United Kingdom after the privatization of Cambridge's highly effective Plant Breeding Institute. After the sale, genomic scientists that had been part of the institute were strongly incentivized by the research-granting councils to focus on research with scientific impact as measured through the rank of the scientific journal where the research is published [20]. With these incentives, most of the wheat genomic scientists shifted their focus away from wheat toward model organism crops such as *Arabidopsis* that are genetically simpler and for which genomic analysis could be generated sooner and published in higher-ranking journals. This diversion of effort left the private wheat research industry without the support of public scientists for nearly 10 years. Recognizing the problem, the principal funder of basic and strategic biological research in the United Kingdom, the Biotechnology and Biological Sciences Research Council (BBSRC), along with other research granting councils, created new programs and incentives for the basic scientist to work with the private breeders, largely resolving the coordination issue [20].

### **2.3 Replication research in the breeding research and innovation process**

The breeding process occurs through a functional interaction of four relatively distinct groups: basic research scientists, applied genomicists, breeders, and farmers. Depending on the institutional (ownership) structure, these interactions can occur within and across at least two dimensions: private and public [21]. Each of the distinct functional groups has a set of goals that can potentially be met by the outputs of other functional groups in the breeding process. These goals shape the incentives for choices of production activities among the groups. The result of the activities of all four groups is the production of agricultural crops that ultimately satisfy consumer demand.

Although the purpose of the plant breeding research and innovation process can be defined as serving the end-users of the crops, each group within the process may have its own immediate goals, which are shaped by the formal or informal metrics used to measure success within each group. The most commonly used metric for determining the performance of academic researchers who tend to focus on basic research rather than on applied research is the number of peer-reviewed publications in top journals

and the number of citations of their published work [22, 23]. In the context of gene mapping, this has generally meant publishing discoveries of new genotype–phenotype associations (QTL), with less interest within the academic community in the subsequent replication and verification of QTL [24–27]. This indicates that the publication of new QTL discoveries represents a key measure of the academic genomicists' performance.

In our view, rather than discovering new QTL, the focus of applied genomicists is primarily on the verification of QTL that can be used for commercial purposes. Only QTL with a certain degree of stability over populations and environments have the potential to be commercially viable. Thus, in addition to the incentives for discovering new QTL, private applied genomicists have incentives to verify the stability of newly discovered QTL. The set of these QTL is established by the academic genomicists who may be seeking QTL that lead to quick publication rather than those associated with commercially desired traits.

Breeders are interested in using the newly discovered QTL for selecting improved breeds and varieties. The goal is to select superior parent lines to create a population with distinct genetic features, which are associated with preferred phenotypic features. This distinct population can then be registered as a new variety or breed.

While profit is the key performance measure for a private breeding program as a whole, the number of new licensed varieties is the main performance measure for public plant breeders. Both private and public plant breeders have incentives to use the results of published QTL studies if they expect these results to be effective means of developing a new variety or a breed. The additional criterion for choosing which varieties will be developed by private breeders is the profitability of the new variety. In cases where the QTL discoveries, reported in peer-reviewed journals, are not sufficiently replicated, breeders need to determine which QTL would be useful for each population anew. Due to the sheer volume of academic QTL publications, the costs of determining which QTL could potentially be useful for private breeders may be high. All this increases the costs to plant breeders of adopting MAS strategies compared to a situation when genotype–phenotype relationships are validated through verification studies.

The costs for plant breeders to adopt MAS may lead to low incentives for using genetic markers as a means of developing new varieties or breeds and thus potentially lower the probability of developing a new variety or breed. Collaboration between geneticists and breeders remains an important challenge [28, 29]. Xu and Crouch state that [30] “high proportion of published markers [are] failing at one or more of the translation steps from research arena to application domain.” Similarly, hold that “MAS has had only a small impact on plant breeding so far.” As one of the reasons for this low impact, they identify the low publishing potential of QTL validation studies:

*“New QTL are frequently reported in scientific journals, but reconfirmation of these QTL in other germplasm and identification of more useful markers are usually not considered novel enough to warrant new publications. This is unfortunate because it is exactly this type of information that is needed for MAS.”*

Ref. [31] also note that the “vast majority of publications on the subject are not considered to have real impact on breeding efforts.”

Thus, the public applied genomicists' incentives to publish new research at the expense of replication or verification studies may not be in alignment with the breeders' needs for reliable genetic information. Note that although MAS has been



adopted in breeding programs for several major crop species, the “majority of the legume crops ... remained untouched with genomics revolution.” Similarly [32], hypothesize that “top journals, concerned with the need to maintain reputations and encourage originality, may be less likely to publish replications.” To assess this hypothesis, we measured the extent of replication research in a sample of genetics journals. We also tested the link between journal rank and the tendency to publish discovery versus replication research.

### **3. Empirical model**

Our investigation of the extent of incentive incompatibility within the plant breeding supply chain consists of two main steps. First is the gathering of data on scientific articles on QTL discovery research and QTL replication research along with the ranking of the journals in which they appeared. Second is the determination of the relationship between the type of QTL research and the journal ranking to test the hypothesis that more prestigious journals publish more articles on MAS discovery rather than verification. The next section of the paper discusses the implication of the results on the design of research programs to ensure incentive compatibility.

Keywords were identified that could be used to search for QTL discovery or replication research in Google scholar. The preliminary search criterion was initially selected through consultations with plant breeding experts. The selection process was supplemented by determining the frequency of keywords and the context in which they were used for the 2019 volume of four genomic journals: Theoretical and Applied Genetics, Euphytica, Journal of Genetics, Genetic Resources and Crop Evolution.

The results from the search of all papers published in 2019 of the four journals are summarized by journal and type of research in **Table 1**.<sup>2</sup> Papers using phrases “QTL Discovery,” “New QTL,” “Novel QTL,” or “Identified QTL” tended to be the most numerous in theoretical and applied genetics and euphytica. Papers using these phrases also tended to be focusing on the identification of particular QTL rather than on general method development for QTL identification. On the other hand, when associated with the words “marker,” “gene or allele,” the attributes “new,” “novel,” “identified,” and “developed” tended to appear in articles that are focusing on developing methods rather than on the identification of particular QTL.<sup>3</sup> When it comes to replication research, “verification,” “replication,” or “confirmation” of a QTL were the phrases most frequently associated with replicating previous QTL research. Based on these findings, we selected “new QTL,” “novel QTL,” and “QTL discovery” as indicators of new QTL research, and “validated QTL,” “confirmed QTL,” “verified QTL,” “QTL validation,” “QTL verification,” and “QTL confirmation” as indicators of replication research.

---

<sup>2</sup> We have prepared a summary of key points for each article that contains keywords denoting new QTL research or replication research. This document is available as supplementary material. **Table 1** is the condensed summary of this document.

<sup>3</sup> When “new,” “novel,” “identified,” and “developed” were associated with markers, genes or allele they focused on the identification of markers, gene or alleles that will help discover future QTL. They would not necessarily outline a complete method to discover QTL but would rather state that this marker will help to discover novel QTL responsible for a certain trait.

Research Type	Keywords used to classify articles	Journal <sup>a</sup>				Primary research focus
		TAG	EUPH	GRCA	JG	
New	Discovery/New/Novel/Identified QTL	64	37	2	2	Identifying particular QTL
	Discovery/New/Novel/Identified/Developed Marker	14	7	15	10	Developing methods
	Novel/Identified gene	20	14	3	4	Developing methods
	Novel/Identified allele	3		6	1	Developing methods
Replication	Verification/Replication/Confirmation of a QTL	22	14	0	2	Identifying particular QTL
	Verification/Replication/Confirmation of a gene	2	7	0	0	Developing methods
	Verification/Replication/Confirmation of marker	0	3	0	3	Developing methods

<sup>a</sup>Journals are *Theoretical and Applied Genetics (TAG)*, *Euphytica (EUPH)*, *Genetic Resources and Crop Evolution (GRCE)*, and *Journal of Genetics (JG)*.

**Table 1.**  
 Frequency of keywords by QTL research type and context for four academic journals in 2019.

A variable was created for each journal to proxy the relative importance of discovery QTL research to overall QTL research published in that journal. Four ratios were calculated to determine the relative importance of discovery versus replication research for each journal. In addition, a fifth ratio was developed to measure the importance of overall QTL research (new or replication-related) in the selected journals. The first three ratios focus on individual keywords associated with new research while the fourth ratio puts all three keywords together. This approach allows us to identify keywords that may be more relevant than others as indicators of differences across journals.

The first ratio,  $Share\ New_i$ , is a measure of the importance of new QTL research relative to overall QTL research for journal  $i$ .

$$ShareNew_i = \frac{New_i}{New_i + Replication_i} \quad (1)$$

where  $New_i$  represent the number of papers in which “new QTL” appeared in journal  $i$ 's online search form<sup>4</sup>, and  $Replication_i$  represents the total number of papers in which “verified QTL,” “confirmed QTL,” or “Validated QTL,” or “QTL verification,” or “QTL confirmation,” or “QTL validation” appeared in journal  $i$ 's online

<sup>4</sup> The journals' online search pages did not have a specified date range. The results were from any volume/issue of the journal that would be available online and on the journal's website.

search form. A second alternative uses the keyword “QTL discovery” to define the following ratio

$$ShareDiscovery_i = \frac{Discovery_i}{Discovery_i + Replication_i} \quad (2)$$

where  $Discovery_i$  represents the number of papers in which “QTL discovery” appeared in journal  $i$ 's online search form. A third uses the keyword novel to represent new QTL research,

$$ShareNovel_i = \frac{Novel_i}{Novel_i + Replication_i} \quad (3)$$

where  $Novel_i$  represents the number of papers in which “Novel QTL” appeared in journal  $i$ 's online search form. Finally, all three keywords were used in the following ratio

$$ShareNewQTL_i = \frac{New_i + Discovery_i + Novel_i}{New_i + Discovery_i + Novel_i + Replication_i} \quad (4)$$

The numerator is the number of papers in which all three keywords indicating new QTL research appeared in a given journal, while the denominator is the number of papers in which all the selected keywords appeared in the same journal. In all four ratios, the higher the value, the greater the importance of marker discovery in the publication of QTL research in that journal. It is important to note that the maximum value of any given ratio is one, which implies that all keywords are associated with new/discovery/novel QTL research. The fifth ratio,  $ShareQTL_i$  represents the share of papers in which any of the selected keywords (both new and replication-related) appeared in the total number of papers published in journal  $i$ :

$$ShareQTL_i = \frac{New_i + Discovery_i + Novel_i + Replication_i}{AllPapers_i} \quad (5)$$

The purpose of this ratio was to assess the overall importance of QTL research across the 24 selected journals and to test for differences in the ratio between higher-ranked journals and lower-ranked journals.

The five ratios defined in Eqs. (1)–(5) were calculated from any volume/issue of a specific journal that was available online and on the journal's website for 24 journals selected in the crop genomics discipline. The ranking of the 24 journals according to their respective average scientific journal rankings (SJR) indicators since 2001 as reported by the SJR (2019) database is listed in **Table 2**. The SJR indicator weighs the influence of a journal, based on the number of citations received by the journal and importance of the journals from which the citations are derived. The higher the SJR indicator, the greater is the prestige of the journal.

The first step in the analysis is to examine the annual frequency of the three new QTL keywords (new, discovery, and novel), the frequency of the six replication QTL keywords (validated, confirmed, verified, validation, verification, and confirmation), and the resulting share measures given by Eqs. (1)–(5). Using the four keywords for QTL discovery research and the five keywords for QTL verification research, Google Scholar was searched to identify refereed articles containing those terms in academic

Rank	Journal	SJR indicator
1	Trends in Genetics	7.49
2	Current Opinion in Plant Biology	5.49
3	PLoS Pathogens	4.56
4	Plant Physiology	3.77
5	Genetics	3.60
6	Evolution	3.39
7	Molecular Ecology	3.15
8	Journal of Experimental Botany	2.26
9	Journal of Evolutionary Biology	2.21
10	BMC genomics	2.00
11	Theoretical and Applied Genetics	2.00
12	Genetics Selection Evolution	1.19
13	Molecular Breeding	1.18
14	BMC Genetics	1.08
15	Journal of Heredity	1.08
16	Crop Science	1.00
17	Cytogenetic and Genome Research	0.95
18	Genome	0.94
19	Tree Genetics & Genomes	0.91
20	Journal of the American Society for Horticultural Science	0.76
21	Euphytica	0.72
22	Journal of Animal Breeding and Genetics	0.69
23	Genetic Resources and Crop Evolution	0.61
24	Journal of Genetics	0.36

**Table 2.**

*Journals used for calculating the relative frequency of keywords referring to new and replication research in genetics and their respective SJR score between 2001 and 2018.*

journals between 2000 and 2019.<sup>5</sup> The average of the numbers of papers containing the respective keywords and the share measures are compared to determine if there is a statistically significant difference between the top 12 ranked journals from the other 12 with lower SJR rankings. In addition, trends are examined to determine if the prominence of the keywords has changed over time.

Finally, a regression analysis is conducted with the five share measures for each journal regressed against its SJR indicator value.

$$Share J_i = \beta_0 + \beta_1 SJR_i + e \quad J = New, Discovery, Novel, New QTL, and QTL \quad (6)$$

<sup>5</sup> This refers to the presence of said phrases anywhere in the paper, not to the keywords listed at the beginning of a paper.

where  $\beta_0$  and  $\beta_1$  are parameters to be estimated. The implied hypothesis is that journal indicators may influence the motivation of the journal to publish new research at the expense of replication research. Alternatively, the causality may run in the opposite direction. It may be that the relative share of new research boosts the ranking of a journal. Either way, there would be a relationship between the quality of the journal and the relative share of new research, and the main aim of our research is to investigate this potential relationship. We hypothesize that a higher SJR indicator is associated with a propensity to publish a greater share of new QTL research, which means that journals with higher SJR indicators would also tend to have higher values for the ratios in Eqs. (1)–(5).

To truly test the claim that top journals have a motivation to publish new research at the expense of replication research, one would need to analyze the rejection rates for new research versus replication research rather than the number of observed published articles. Unfortunately, the data on rejection rates are not available. This is our best attempt to get around this data problem. The downside is that the strength of our conclusions is reduced. Thus, rather than testing the hypothesis of journal bias in favor of new research over replication research, we are assessing potential links and relationships.

#### 4. Results

The number of papers containing the selected keywords related to QTL research that are available in the Google Scholar database for all journals from the years 2000 through 2019 are listed in **Table 3**. The frequency of papers with QTL described as *New* or *Discovery* or *Novel* has increased steadily since 2000. The trend variables suggest that each of the research articles containing the key search words “*New*” and “*Discovery*” have increased annually by approximately 11 papers per year while the number of papers with “*Novel*” has increased by nearly 17 papers per year. Articles with keywords for replication research have also increased over time, but the absolute rate of annual increase is significantly less than the increase in papers containing new QTL keywords.

Year	Papers with new QTL keywords			Papers with replication QTL keywords		New QTL as share of total measures			
	New	Discovery	Novel	Verified or confirmed or validated	Verification or confirmation or validation	Share new	Share discovery	Share novel	Share new QTL
2000	40	4	14	5	7	0.77	0.25	0.54	0.83
2001	42	5	8	8	10	0.70	0.22	0.31	0.75
2002	51	12	17	8	29	0.58	0.24	0.31	0.68
2003	68	19	23	14	37	0.57	0.27	0.31	0.68
2004	92	42	38	18	33	0.64	0.45	0.43	0.77
2005	91	43	53	29	37	0.58	0.39	0.45	0.74
2006	135	34	70	13	52	0.68	0.34	0.52	0.79
2007	119	57	69	18	61	0.60	0.42	0.47	0.76
2008	131	60	76	29	73	0.56	0.37	0.43	0.72

Year	Papers with new QTL keywords			Papers with replication QTL keywords		New QTL as share of total measures			
	New	Discovery	Novel	Verified or confirmed or validated	Verification or confirmation or validation	Share new	Share discovery	Share novel	Share new QTL
2009	158	83	94	41	83	0.56	0.40	0.43	0.73
2010	135	78	106	21	104	0.52	0.38	0.46	0.72
2011	159	89	121	34	111	0.52	0.38	0.45	0.72
2012	198	118	126	35	119	0.56	0.43	0.45	0.74
2013	203	108	139	37	92	0.61	0.46	0.52	0.78
2014	199	114	184	41	120	0.55	0.41	0.53	0.76
2015	211	128	187	38	138	0.55	0.42	0.52	0.75
2016	218	137	226	34	117	0.59	0.48	0.60	0.79
2017	204	174	265	41	129	0.55	0.51	0.61	0.79
2018	226	208	303	39	128	0.58	0.55	0.64	0.82
2019	257	220	403	38	133	0.60	0.56	0.70	0.84
Mean	146.85	86.65	126.1	27.05	80.65	0.59	0.40	0.48	0.76
Trend	11.06*	10.53*	16.92*	1.86*	7.18*	-0.006	0.014*	0.014*	0.003
Std Error	0.54	0.63	1.49	0.24	0.46	0.002	0.001	0.002	0.001

\*Coefficient on trend variable is statistically significant at the 99% confidence level.

**Table 3.**  
 Frequency of QTL research papers Google Scholar from 2000 to 2019.

Although the rate of increase is higher for the keywords related to new QTL discovery compared to the replication QTL keywords, the relative focus on new QTL research compared to replication research depends on the choice of keywords. The *Share New* and *Share New QTL* ratios did not change significantly over time (**Table 3**). However, the value of the *Share Discovery* and *Share Novel* increased by approximately 1.4 percentage points annually over the last two decades.

In order to assess whether the increasing focus on new QTL research versus replication has been driven in part by incentives for academic genomicists, we examined the frequency of the keywords for each of the 24 journals in **Table 4**. The selected journals in **Table 4** are listed in the order of their SJR ranking (see **Table 2**) and the number of articles, in which the keywords associated with new QTL research and replication QTL research, are given along with the four ratios indicating the share of new QTL research to total QTL research. The averages of these measures are calculated for the top 12 ranked and the bottom 12 of the journals chosen for analysis and a t-test<sup>6</sup> used to determine if the difference in averages is statistically significant.

The top 12 journals tended to have higher values for “*New*,” “*Discovery*,” and “*Novel*” as hypothesized as compared to the bottom 12 of the journals selected. There is no difference in the average appearance of keywords related to replication between

<sup>6</sup> The test was paired, two samples for means.

Journal	New QTL keywords			Replication QTL keywords		Total publications	New QTL as share of total measures				Share QTL
	New	Discovery	Novel	Verified or confirmed or validated	Verification or confirmation or validation		Share new	Share discovery	Share novel	Share new QTL	
Trends in Genetics	49	6	1	0	0	6650	1.00	1.00	1.00	1.00	0.0084
Current Opinion in Plant Biology	2	5	4	0	3	2240	0.40	0.63	0.57	0.79	0.0063
PLoS Pathogens	2	1	3	0	1	3180	0.67	0.50	0.75	0.86	0.0022
Plant Physiology	7	10	7	2	7	102,000	0.44	0.53	0.44	0.73	0.0003
Genetics	79	23	23	21	23	121,000	0.64	0.34	0.34	0.74	0.0014
Evolution	5	1	1	0	0	102,000	1.00	1.00	1.00	1.00	0.0001
Molecular Ecology	6	6	5	0	2	15,100	0.75	0.75	0.71	0.89	0.0013
Journal of Experimental Botany	29	14	37	5	27	16,100	0.48	0.30	0.54	0.71	0.0070
Journal of Evolutionary Biology	0	1	0	0	0	4640	0.00	1.00	0.00	1.00	0.0002
BMC Genomics	36	23	34	13	16	14,500	0.55	0.44	0.54	0.76	0.0084
Theoretical and Applied Genetics	177	54	129	83	100	12,700	0.49	0.23	0.41	0.66	0.0428
Genetics Selection Evolution	21	6	15	7	2	1890	0.70	0.40	0.63	0.82	0.0270
Average for the top 12 journals based on SJR Indicator	34.42	12.50	21.58	10.92	15.08	33,500	0.59	0.59	0.58	0.83	0.0088
Molecular Breeding	70	34	56	43	60	5580	0.40	0.25	0.35	0.61	0.0471
BMC Genetics	23	8	19	14	8	4130	0.51	0.27	0.46	0.69	0.0174
Journal of Heredity	14	1	3	3	3	8150	0.70	0.14	0.33	0.75	0.0029
Crop Science	66	39	45	32	29	70,700	0.52	0.39	0.42	0.71	0.0030
Cytogenetic and Genome Research	2	0	0	1	0	6180	0.67	0.00	0.00	0.67	0.0005

Journal	New QTL keywords			Replication QTL keywords		Total publications	New QTL as share of total measures				Share QTL
	New	Discovery	Novel	Verified or confirmed or validated	Verification or confirmation or validation		Share new	Share discovery	Share novel	Share new QTL	
Genome	9	3	13	0	0	5867	1.00	1.00	1.00	1.00	0.0043
Tree Genetics & Genomes	9	12	1	4	14	2080	0.33	0.40	0.05	0.55	0.0192
Journal of the American Society for Horticultural Science	1	2	4	1	1	4140	0.33	0.50	0.67	0.78	0.0022
Euphytica	82	16	43	29	31	9860	0.58	0.21	0.42	0.70	0.0204
Journal of Animal Breeding and Genetics	10	1	2	2	6	1860	0.56	0.11	0.20	0.62	0.0113
Genetic Resources and Crop Evolution	0	2	4	2	1	2860	0.00	0.40	0.57	0.67	0.0031
Journal of Genetics	6	0	4	1	1	81,900	0.75	0.00	0.67	0.83	0.0001
Average for the bottom 12 journals based on SJR Indicator	24.33	9.83	16.17	11.00	12.83	16,942	0.53	0.31	0.43	0.71	0.0110
difference in mean between top 12 and bottom 12	10.09	2.67	5.40	-0.08	2.25	16,557	0.06	0.28*	0.15*	0.12*	0.0022
t-test statistic	0.52	0.41	0.40	-0.01	0.20	1.15	0.69	3.22	1.54	2.65	

\*Difference in averages is statistically significant at the 90% confidence level.

**Table 4.** Frequency of Papers with QTL research keywords in each of 24 journals by SJR ranking from 2000 to 2019.



the two groups of journals categorized by SJR ranking. Given the greater appearance of new QTL keywords in the top 12 ranked journals compared to the other 12 and the insignificant difference QTL replication keywords between the two groups of journals, the relative role of new QTL as a share of total QTL research is higher in the top-ranked journals is expected. The difference is particularly high for the *Share Discovery* ratio and statistically insignificant for *Share New*.

Next, the difference in the share of papers mentioning QTL research (new or replication) between the top 12 and the bottom 12 journals was not statistically significant. This indicates that there was no significant difference in the level of emphasis on overall QTL research between the two groups. This suggests that, even though both groups of journals publish QTL research, the top group puts more emphasis on new QTL research while the bottom group puts more focus on QTL replication research. The total number of papers published was higher in the top 12 journals. This might suggest that QTL research in the top journals competes with a greater number of topics than in the bottom 12 journals. Higher-ranked journals may publish cutting-edge lines of research that are not present in the lower-ranked papers. In this setting, replication QTL research may be at a competitive disadvantage when competing not only against new QTL research but also against other advanced lines of research in top-ranked journals.

A final step in the empirical analysis is to examine the relationship between a journal's SJR indicator and the value of the four ratios used to proxy the relative importance of new QTL research to overall QTL research in the journal. The results of the five regressions (Eq. 5) are listed in **Table 5**. There is a positive relationship between the focus on new QTL discovery and journal rank as defined by its SJR indicator as hypothesized for the first four ratios. The fifth ratio, measuring overall QTL research, had a negative but not statistically significant coefficient on Share QTL. The intercepts were positively significant at a 99% confidence level for all five models, which is expected as the SJR indicator is generally a number greater than zero for most journals. The slope for *Share Discovery* was positive and significant at the 99% confidence level, while the slopes for *Share Novel* and *Share New QTL* were positive and significant at the 95% confidence level. This result is consistent with the hypothesis

Dependent variable	Intercept	SJR indicator	R <sup>2</sup>	Adjusted R <sup>2</sup>
Share New	0.4721** (0.0812) <sup>a</sup>	0.0416 (0.0416)	0.0841	0.042
Share Discovery	0.2461** (0.0841)	0.095** (0.0304)	0.308	0.276
Share Novel	0.3656** (0.0846)	0.0643* (0.0305)	0.168	0.130
Share New QTL	0.6991** (0.0378)	0.0344* (0.0136)	0.2245	0.189
Share QTL	0.013032** (0.0042)	-0.0015 (0.0015)	0.0411	-0.002

<sup>a</sup>Standard errors are in parentheses. \*\*Significant at the 99% confidence level. \*Significant at 95% confidence level.

**Table 5.** Regression ( $n = 24$ ) results between focus on new discovery research and SJR indicator (Eq. 5).

that more highly ranked journals tend to favor new research over replication or verification research, while the extent of overall QTL research is not significantly affected by journal rank.

## 5. Conclusions

The purpose of this paper was to develop a better understanding of incentives for research and innovation within the plant breeding process. Advances in genomic technology have brought the potential for significant gains in agricultural productivity within a much shorter time frame than possible with traditional phenotypic breeding strategies. The breeding supply chain has expanded with genomic technology to include basic research scientists and applied genomicists along with traditional plant breeders and farmers. Capturing the gains made possible by genomic technology will require cooperation through the key stakeholders within this plant breeding supply chain.

Genomicists have focused on identifying specific DNA sequences or QTLs that can be used as molecular markers. However, the use of MAS by breeders in their programs requires the identified QTL to be reliably correlated with agronomically desired traits. Replication research is critical for reducing the risk associated with the adoption of new marker-based (or QTL-based) selection strategies, but the applied scientists doing genomics research often do not have the incentives to do replication and other research required to verify the reliability of markers. The misalignment of incentives in the breeding supply chain can curtail the development of the projected advances in food production by genomics research.

The metric used to reward basic and applied genomic scientists is the prestige or higher rank of the journal where their research is published, and this has created a bias toward identifying new markers rather than the verification of existing markers. Using a sample of 24 genomic journals, we found more highly ranked journals tend to favor new research on identifying new QTL over replication research on previously identified QTL. Given that, breeders will tend to adopt only those markers perceived to be reliable, the implicit lack of incentives for basic and applied genomic scientists to undertake replication research can impede agricultural innovation.

However, there may be other factors influencing academic geneticists' decision to perform replication research. Rajsic et al. [33] find that cost considerations are important in determining the sizes of training populations and the number of replications. QTL validation examines whether the same QTL appears when the genetic background is grown in other locations and/or years and whether its effect can still be detectable when introduced into a different genetic background. Lack of appropriate funds for public research may contribute to an overall lack of replication studies published by academic geneticists. Although the lack of replication done by academic geneticists could explain an overall low ratio of replication research to new research, it is hard to see why this would cause differences between top-ranked journals and lower-ranked journals. This suggests that incentives play a role in addition to other potential factors.

Policymakers designing breeding research and innovation programs must recognize the potential for misalignment of incentives within the supply chain. Rather than reward applied scientific researchers on the basis of publication surrounding new QTL in high-impact journals, funding agencies should create incentives for basic scientists to work with the breeders to focus on the identification and replication of traits

desired at the farm level. Alternatively, funders of large research projects targeted toward a variety of development could require, and/or fund, additional verification studies for new QTL. In the absence of policy change, the lack of verification will continue to be an impediment to crop innovation.

## **Acknowledgements**

Funding for the research was partially provided by the Applied Bean Genomics and Bioproducts Project, a multi-interinstitutional partnership between the University of Guelph, the University of Western Ontario, the University of Windsor, and Agriculture and Agri-Food Canada, Génome Québec & Genome Canada SoyaGen.

## **Author details**

Predrag Rajsic<sup>1\*</sup>, Richard Gray<sup>2</sup>, Alfons Weersink<sup>3</sup> and Istvan Rajcan<sup>4</sup>

1 Department of Economics, University of Waterloo, Ontario, Canada


2 Department of Agricultural and Resource Economics, University of Saskatchewan, Saskatoon, Saskatchewan, Canada

3 Department of Food, Agricultural and Resource Economics, University of Guelph, Guelph, Ontario, Canada

4 Department of Plant Agriculture, University of Guelph, Guelph, Ontario, Canada

\*Address all correspondence to: [prajsic@uwaterloo.ca](mailto:prajsic@uwaterloo.ca)

## **IntechOpen**

© 2023 The Author(s). Licensee IntechOpen. This chapter is distributed under the terms of the Creative Commons Attribution License (<http://creativecommons.org/licenses/by/3.0>), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited. 

## References

- [1] Fróna D, Szenderák J, Harangi-Rákos M. The challenge of feeding the world. *Sustainability*. 2019;**11**(20):5816
- [2] Valoppi F, Agustin M, Abik F, Morais de Carvalho D, Sithole J, Bhattarai M, et al. Insight on current advances in food science and technology for feeding the world population. *Frontiers in Sustainable Food Systems*. 2021;**5**:626227
- [3] Fuglie KO, Wang SL. New Evidence Points to Robust but Uneven Productivity Growth in Global Agriculture. *Global Journal of Emerging Market Economies*. 2013;**5**:23-30
- [4] OECD. *Innovation, Productivity and Sustainability in Food and Agriculture: Main Findings from Country Reviews and Policy Lessons*. Paris: OECD Publishing; 2019
- [5] Alston JM, Beddow JM, Pardey PG. Agricultural research, productivity, and food prices in the long run. *Science*. 2009;**325**(5945):1209-1210
- [6] Kafatos FC, Eisner T. Unification in the century of biology. *Science*. 2004;**303**(5662):1257-1257
- [7] Warr A, Affara N, Aken B, Beiki H, Bickhart DM, Billis K, et al. An improved pig reference genome sequence to enable pig genetics and genomics research. *Gigascience*. 2020;**9**(6):giaa051
- [8] Love B. Personal communication. (P. Rajsic, ed.). 2013
- [9] Ioannidis JP. Differentiating Biases from Genuine Heterogeneity: Distinguishing Artifactual from Substantive Effects. *Publication bias in Meta-analysis: Prevention, Assessment and Adjustments*. Chichester, West Sussex, England. 2005. pp. 287-302
- [10] Open Science Collaboration. Estimating the reproducibility of psychological science. *Science*. 2015;**349**:6251
- [11] Baker M. Reproducibility crisis. *Nature*. 2016;**533**(26):353-366
- [12] Fanelli D. Opinion: Is science really facing a reproducibility crisis, and do we need it to? *Proceedings of the National Academy of Sciences*. 2018;**115**(11):2628-2631
- [13] Rasheed A, Hao Y, Xia X, Khan A, Xu Y, Varshney RK, et al. *Crop Breeding Chips and Genotyping Platforms: Progress, Challenges, and Perspectives*. *Molecular Plant*. 2017;**10**:1047-1064
- [14] Porter, M. E. (2011). *Competitive advantage of nations: creating and sustaining superior performance*. simon and schuster
- [15] Klerkx L, van Mierlo B, Leeuwis C. Evolution of systems approaches to agricultural innovation: concepts, analysis and interventions. In: Darnhofer I, Gibbon D, Dedieu B, editors. *Farming Systems Research into the 21st Century: The New Dynamic*. Netherlands, Dordrecht: Springer; 2012. pp. 457-483
- [16] Spielman DJ, von Grebmer K. *Public-private Partnerships in Agricultural Research: An Analysis of Challenges Facing Industry and the Consultative Group on International Agricultural Research*. Washington DC: International Food Policy Research Institute (IFPRI); 2004
- [17] Heisey PW, Fuglie KO. Public agricultural R&D in high-income countries: Old and new roles in a new funding environment. *Global Food Security*. 2018;**17**:92-102

- [18] Gray RS, Kingwell RS, Galushko V, Katarzyna B. Intellectual Property Rights and Canadian Wheat Breeding for the 21st Century. *Canadian Journal of Agricultural Economics/Revue canadienne d'agroeconomie*. 2017;**65**:667-691
- [19] Varshney RK, Kudapa H, Roorkiwal M, Thudi M, Pandey MK, Saxena RK, et al. Advances in genetics and molecular breeding of three legume crops of semi-arid tropics using next-generation sequencing and high-throughput genotyping technologies. *Journal of Biosciences*. 2012;**37**:811-820
- [20] Galushko V, Gray R. Twenty five years of private wheat breeding in the UK: Lessons for other countries. *Science and Public Policy*. 2014;**41**:765-779
- [21] Morris M, Edmeades G, Pehu E. The global need for plant breeding capacity: what roles for the public and private sectors? *HortScience*. 2006;**41**:30-39
- [22] Oppenheim C. The correlation between citation counts and the 1992 research assessment exercise ratings for British research in genetics, anatomy and archaeology. *Journal of Documentation*. 1997;**53**:477-487
- [23] Li L, Zhang H. Confidentiality and information sharing in supply chain coordination. *Management Science*. 2008;**54**:1467-1481
- [24] Ioannidis JP, Ntzani EE, Trikalinos TA, Contopoulos-Ioannidis DG. Replication validity of genetic association studies. *Nature Genetics*. 2001;**29**:306-309
- [25] Lohmueller KE, Pearce CL, Pike M, Lander ES, Hirschhorn JN. Meta-analysis of genetic association studies supports a contribution of common variants to susceptibility to common disease. *Nature Genetics*. 2003;**33**:177-182
- [26] NCI-NHGRI Working Group on Replication in Association Studies. Replicating genotype–phenotype associations: what constitutes replication of a genotype–phenotype association, and how best can it be achieved? *Nature*. 2007;**447**:655-660
- [27] Hirschhorn JN, Lohmueller K, Byrne E, Hirschhorn K. A comprehensive review of genetic association studies. *Genetics in Medicine*. 2002;**4**:45-61
- [28] Collard BC, Mackill DJ. Marker-assisted selection: an approach for precision plant breeding in the twenty-first century. *Philosophical Transactions of the Royal Society, B: Biological Sciences*. 2008;**363**:557-572
- [29] de Dorlodot S, Forster B, Pagès L, Price A, Tuberosa R, Draye X. Root system architecture: opportunities and constraints for genetic improvement of crops. *Trends in Plant Science*. 2007;**12**:474-481
- [30] Xu Y, Crouch JH. Marker-assisted selection in plant breeding: from publications to practice. *Crop Science*. 2008;**48**:391-407
- [31] Scheben A, Batley J, Edwards D. Genotyping-by-sequencing approaches to characterize crop genomes: choosing the right tool for the right application. *Plant Biotechnology Journal*. 2017;**15**:149-161
- [32] Singh K, Ang SH, Leong SM. Increasing Replication for Knowledge Accumulation in Strategy Research. *Journal of Management*. 2003;**29**:533-549
- [33] Rajsic P, Weersink A, Navabi A, Pauls KP. Economics of genomic selection: the role of prediction accuracy and relative genotyping costs. *Euphytica*. 2016;**2**:259-276